9-2013

Development of Mass Spectrometry-Based Methods for Quantitation and Characterization of Protein Drugs: Transferrin as a Model Drug Delivery Vehicle

Shunhai Wang
University of Massachusetts Amherst, shunhaiwang@gmail.com

Development of Mass Spectrometry-Based Methods for Quantitation and Characterization of Protein Drugs: Transferrin as a Model Drug Delivery Vehicle

A Dissertation Presented

by

SHUNHAI WANG

Submitted to the graduate school of the

University of Massachusetts Amherst in partial fulfillment

of the requirement for the degree of

DOCTOR OF PHILOSOPHY

SEPTEMBER 2013

Department of Chemistry

DEVELOPMENT OF MASS SPECTROMETRY-BASED METHODS FOR QUANTITATION AND CHARACTERIZATION OF PROTEIN DRUGS: TRANSFERRIN AS A MODEL DRUG DELIVERY VEHICLE

A Dissertation Presented

by

SHUNHAI WANG

Approved as to style and content by:

_____

Igor A. Kaltashov, Chair

_____

Richard W. Vachet, Member

_____

Anne B. Mason, Member

_____

Stephen J. Eyles, Member

_____

Ricardo B. Metz, Acting Department Head
Department of Chemistry

DEDICATION

This dissertation is dedicated to my father, Changshou Wang and my mother Bingfang Yao. Their support, encouragement and constant love have sustained me throughout my life.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Igor A. Kaltashov, not only for many years of professional guidance on my scientific training, but also for his encouraging and persistent support. His supervision during the five years is invaluable for my growth in both professionalism and character. Thanks are also due to Cedric E. Bobst and Rinat R. Abzalimov. The discussions with them have always been inspiring. I also wish to express my appreciation to the rest of the group members, Adriana Kita, Burcu Baykal, Guanbo Wang, Nguyen Son, Hanwei Zhao, Jake Pawlowski, Khaja Muneeruddin, Shengsheng Xu and Ololade Fatunmbi. All of them have made this group a great place to study and work. And their friendship will be an invaluable fortune for my entire life.

I would also like to express my gratitude to the members of my committee, Richard Vachet, Anne B. Mason, and Stephen Eyles, for their helpful comments and suggestions on all stages of my work. A special thank you to Anne B. Mason for generously providing transferrin samples for my study.

I also want to thank Gokhan Elci, Bo Yan and Yuqing Xing from Prof. Vachet's group for ICP-MS measurement. Furthermore, I want to express my gratitude to Dr. Sung Tae Kim, Krishnendu Saha and Dr. Chang Soo Kim from Prof. Rotello's group for the animal work.

I also want to thank National Institute of Health (R01 GM061666) for funding this research, and Graduate School and Chemistry Department of University of Massachusetts for providing some travel expenses during conferences.

At last, a special thank you to all my friends and family members. It is your support and encouragement that I could stay focused on this dissertation work.

ABSTRACT

DEVELOPMENT OF MASS SPECTROMETRY-BASED METHODS FOR QUANTITATION AND CHARACTERIZATION OF PROTEIN DRUGS: TRANSFERRIN AS A MODEL DRUG DELIVERY VEHICLE

SEPTEMBER 2013

SHUNHAI WANG, B.S., PEKING UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Igor A. Kaltashov

In the last two decades, protein drugs have enjoyed a rapid growth and achieved a tremendous success in treating human diseases. However, the presence of physiological barriers greatly impedes the efficient delivery of such unconventional large molecule drugs, and therefore limits their clinical utility. An elegant way to address this challenge takes advantage of certain endogenous transporter proteins, such as human transferrin (Tf), whose ability to traverse physiological barriers has been extensively exploited. However, methods to investigate Tf-based drug delivery remained insufficient and unsatisfactory until recent development of quantitative mass spectrometry (MS). Hereby, MS-based methods have been developed and validated for quantitation of exogenous Tf in biological fluids. Particularly, different O18-labeling based approaches have been evaluated, modified and developed in this work, in order to achieve the most reliable quantitation. Alternatively, a novel approach based on indium labeling and inductively coupled plasma mass spectrometry (ICP-MS) detection has been developed for sensitive

quantitation of Tf in biological fluids. The second aspect of this dissertation work focuses on the application of MS-based methods for characterization of protein drugs at different levels, ranging from protein identification, covalent structure, conformation, and interaction with physiological partners. Particularly, an O18-labeling assisted approach has been developed to identification of protein deamidation products. This new approach can readily distinguish between the two deamidated isomers. Also, an LC-MS based method has been developed for ranking the susceptibility of protein disulfide bonds to reduction, which could be applied to several disulfide bond-related analyses. Finally, a recently designed growth hormone transferrin fusion protein was studied using MS-based methods, and the molecular basis for its successful oral delivery was revealed.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Protein drugs: development and challenges

High affinity and exquisite specificity towards the therapeutic targets are two of the most important features of an ideal drug. Indeed, those two features have led to a spectacular success of protein-based therapeutics in combating human diseases in the past decades. The first protein drug, biosynthetic "human" insulin was approved to therapeutic use in 1982. Twenty years later, protein-based therapeutics have already become one of the fastest growing segments among the new pharmaceutical products, which representing 44% of all drugs in the development phase and 27% of all drugs in both preclinical and clinical trials [1]. It is expected that protein-based therapeutics will continue to enjoy this growth, and gradually will outpace the traditional small molecule pharmaceuticals. In spite of the success, the clinical utility of many protein drugs is limited due to the inability to efficient and accurate delivery. A significant challenge faced by most protein drugs is the presence of various physiological barriers. For example, intracellular delivery of drugs is very difficult to achieve because the passive entry across the plasma membrane is restricted to molecules that are sufficiently non-polar and less than 500 Da in size [2]. However, none of the protein drugs fall into this category. In addition, when targeting central nervous system (CNS) is necessary, the existence of blood-brain barrier (BBB) amplifies the challenge. Consisting of tightly linked

cerebrovascular endothelial cells, the BBB presents a diffusional barrier between circulating blood and the interstitial space of the brain, making the brain inaccessible to large or hydrophilic molecules by diffusion. In fact, all the protein drugs and more than 98% of small-molecule drugs cannot cross the BBB passively [3]. Additionally, another physiological barrier that presents a formidable challenge in developing orally deliverable protein drugs is the intestinal epithelial barrier. Similar to the BBB, it tightly controls the passage of nutrients and chemicals, as well as protein drugs. In addition to these physiological barriers, the presence of intracellular and extracellular enzymes forms enzymatic barriers that further amplify the challenge of efficient delivery of protein-based therapeutics.

In spite of the difficulty, it has been long recognized that it is essential to develop a molecular mechanism to achieve efficient and accurate delivery of protein drugs [4]. An elegant strategy takes advantage of the ability of certain endogenous proteins to cross physiological barriers via receptor-mediated endocytosis/transcytosis. Those carrier proteins can be utilized as drug delivery vehicle, while the therapeutic proteins can be tethered to them as payload by either conjugation or as recombinant fusion proteins.

1.2 Transferrin as a drug delivery vehicle

Serum transferrin (Tf) is one such carrier protein whose ability to enter cells and cross the physiological barriers via transferrin receptor (TfR) mediated endocytosis/transcytosis has been extensively exploited. Several features have made Tf a

versatile drug delivery vehicle. First, Tf is one of the few proteins that can be internalized by cells via TfR-mediated endocytosis. The pathway of endocytosis has been extensively studied. Fe_2Tf binds the receptor at the cell surface and both are internalized and transported to endosomes. ATP-dependent proton pumps then force hydrogen ions into the endosome reducing the pH to 5.5, which facilitates a conformational change in Tf, and its subsequent release of iron. Then the Apo-Tf/TfR complex returns to the cell surface where Apo-Tf is displaced by Fe2Tf. Apo-Tf molecule then circulates until it again comes in contact with free iron, and the cycle is continued [5]. Importantly, overexpression of TfR is normally observed at the surface of rapidly growing malignant cells due to their elevated iron demands [6], providing the basis for targeted delivery of anti-cancer drugs to tumors (Figure 1-1, left panel). Furthermore, since TfR is one of the abundant proteins in the BBB, it is not surprising that Tf has been a focus of extensive efforts to target central nervous system via TfR-mediated transcytosis (Figure 1-1, right panel). In addition, owing to the wide distribution of the TfR on the inner lining of GI tract [7], Tf-based protein drugs have also been exploited for oral administration [8], which is often seen as the "holy grail" of protein therapy [9, 10].

1.3 Tf-based drug system under development

A variety of therapeutic agents (chemotherapeutic drugs, protein toxins, radionuclides, liposomes, modified viral particles, and nanoparticles) have been tethered to Tf, and studied for delivery to malignant cells by targeting TfR on the cell

surface (Figure 1-2) [11]. For example, Hartinger *et al.* [12] recently reported a Ru-based anticancer agent termed KP1019 (Figure 1-3), which has already entered clinical trials. Taking advantage of its ability to form a covalent adduct with Tf, targeted delivery of this small molecule drug to tumor cells is enhanced. In another example, after conjugating to Tf, artemisinin can be delivered into tumor cells and consequently kill them, by producing free radicals from the reaction with iron [13]. Due to the elevated iron uptake by tumor cells, both toxicity and selectivity of this agent are enhanced. In addition to small molecule drugs, Tf has also been exploited as the delivery vehicle for various protein toxins, such as gelonin [14] and ricin [15]. Notably, CRM 107, a point mutant of diphtheria toxin (DT) with reduced non-specific binding to mammalian cells, has been conjugated to Tf and investigated as a treatment of malignant gliomas [16]. This is the only Tf-cytotoxin conjugate that has reached clinical trials in recent years. Furthermore, high molecular weight compounds, including liposome [17], viruses [18], and nanoparticles [19] have also been conjugated to Tf, and applied for targeted delivery to malignant cells. In addition to Tf, other proteins that bind to TfR have also been exploited as an alternative approach to target tumor cells, including anti-TfR monoclonal antibodies (mAbs) and Tf oligomers [20] (Figure 1-2).

As one of the very few proteins that can traverse the BBB, Tf has been a focus of extensive efforts to deliver therapeutic agents to CNS. In a recent study, human serum albumin nanoparticles which can be loaded with different drugs were conjugated to transferrin, and successful BBB-crossing was observed in an animal mode study [21].

Tf has also been exploited for the design of orally deliverable protein drugs. The wide distribution of TfR on the inner lining of the GI tract [7] provides a possibility of using TfR-mediated transcytosis to overcome the intestinal-epithelial barrier. Applying this strategy, Shen and co-workers [8] recently designed a human growth hormone/Tf fusion protein (GHT) that represented one of the very few examples of a protein drug capable of exhibiting therapeutic response after oral administration. Analysis of this fusion protein in our group [22] revealed the important role of protein aggregation to survive the harsh GI tract environment, and the key role of Tf-moiety to overcome the intestinal-epithelial barrier.

1.4 Challenges for Tf-based drug delivery

Despite the success of conjugating various therapeutic agents to Tf, their efficient delivery is challenging. It is well known that during the receptor-mediated endocytosis, Tf molecule always stays in the endosome while inside the cell, and is quickly recycled back to the cell surface (the entire cycle is only 4 to 5 min) [23]. In order to achieve efficient delivery, the drug molecule must dissociate from Tf and cross the endosomal membrane soon enough to avoid returning to the circulation. It is difficult to meet both requirements. For example, for every 10 million Tf/gelonin cytotoxin conjugates that are recycled, only one molecule of gelonin is actually delivered into the cell in a single cycle [14]. Apparently, the rapid recycling feature of Tf is undesirable for an efficient drug delivery vehicle. Thus, a lot of effort has been made to improve the delivery efficiency by

engineering Tf variants. In a recent study, Kamei *et al* [24] have applied a mathematical model of the Tf trafficking pathway in order to identify the key factors determining the cellular association of Tf, which has already been demonstrated to be correlated with the effectiveness of drug delivery [14]. In this study, they first identified that reduced iron release rate of holo-Tf can effectively increase its cellular association. By replacing the synergistic carbonate anion with oxalate [24], or using a genetically engineered Tf mutant (K296E/K534A) [23] that both led to higher affinity of Tf to iron and slower iron release, they observed a significant increase in Tf uptake by HeLa cells. In addition, application of Tf oligomers is another strategy to alter the Tf trafficking pathway in order to improve the drug delivery efficiency [20].

1.5 Methods to investigate Tf-based drug delivery

It thus is essential to investigate the ability of Tf and its variants as drug delivery vehicle for both cellular uptake and traversing the physiological barriers. To achieve this goal, accurate and sensitive quantitation of exogenous Tf in complex biological fluids is required. Currently, immunoaffinity-based assays are still the mainstay of detection and quantitation of proteins in complex matrices due to the high sensitivity and specificity. However, in the case of Tf quantitation, this well-established technique is less useful due to the presence of abundant endogenous Tf molecules in most biological fluids. For example, in a model study where quantitation of spiked human Tf (hTf) in bovine serum is required, the application of immunoaffinity-based method could be hampered due to

the cross-reactivity of anti-hTf antibody to abundant bovine Tf (bTf). Indeed, those two molecules share ca. 69% of sequence identity and even higher similarity, and therefore it will be an extremely challenging, time-consuming and resource-intensive effort to produce the antibody that can discriminate between those two molecules.

In fact, current methods used to study Tf-based drug delivery mostly rely on different labeling approaches. For instance, taking advantage of the binding ability of Tf to certain metal ions, radioactive labeling technique can be applied by incorporating radioactive metal ions such as ^{59}Fe, ^{67}Ga and ^{111}In [25] into the Tf binding sites. The latter two are frequently used in medical and diagnostic applications [26-28]. In addition, radioactive-iodine labeling (usually ^{125}I) is also frequently applied where the aromatic rings of tyrosine side chains are primarily labeled. Using this approach, detection and quantitation of Tf can be achieved with extremely high sensitivity by measuring the radioactivity [24]. However, due to the cost and safety concerns associated with handling radioactive compounds, the routine use of this technique is limited [29]. In addition, in the case of Tf labeling with radioactive metal ions, the quantitation results are less useful after the ion release during the endocytosis/transcytosis. Furthermore, this approach is not suitable for multiplexing so that two or more different forms of exogenous Tf cannot be investigated in a single experiment. Fluorophore labeling is another strategy to discriminate and quantitate exogenous Tf in the context of abundant endogenous Tf. A great advantage provided by this technique is the imaging capability, so that the bio-distribution of Tf can be studied [30]. In addition, applying this

strategy, different forms of exogenous Tf can be achieved simultaneously by using different chromophores. However, for nearly all the labeling-based strategies, quantitation of Tf is achieved indirectly by measuring the signal from reporter groups (radioactive elements or fluorophores), and therefore labeling efficiency and stability become great concerns.

In the 2nd chapter of this dissertation, novel methods based on mass spectrometry and stable isotope labeling were developed and validated for quantitation of exogenous Tf in biological fluids. Particularly, different O18-labeling based approaches have been evaluated, modified and developed in order to achieve the most reliable quantitation. Alternatively, a novel approach based on indium labeling and inductively coupled plasma mass spectrometry (ICP-MS) detection has been developed for sensitive quantitation of Tf in biological fluids. The 3rd chapter of this dissertation discussed the development of an O18-labeling assisted approach to identification of protein deamidation products. This new approach can readily distinguish between the two deamidated isomers. In the 4th chapter of this dissertation work, we discussed the development of an LC-MS based method for ranking the susceptibility of protein disulfide bonds to reduction, which could be applied to several disulfide bond-related analyses. Finally, in the 5th chapter, we studied a recently designed growth hormone transferrin fusion protein, and the molecular basis for its successful oral delivery was elucidated.

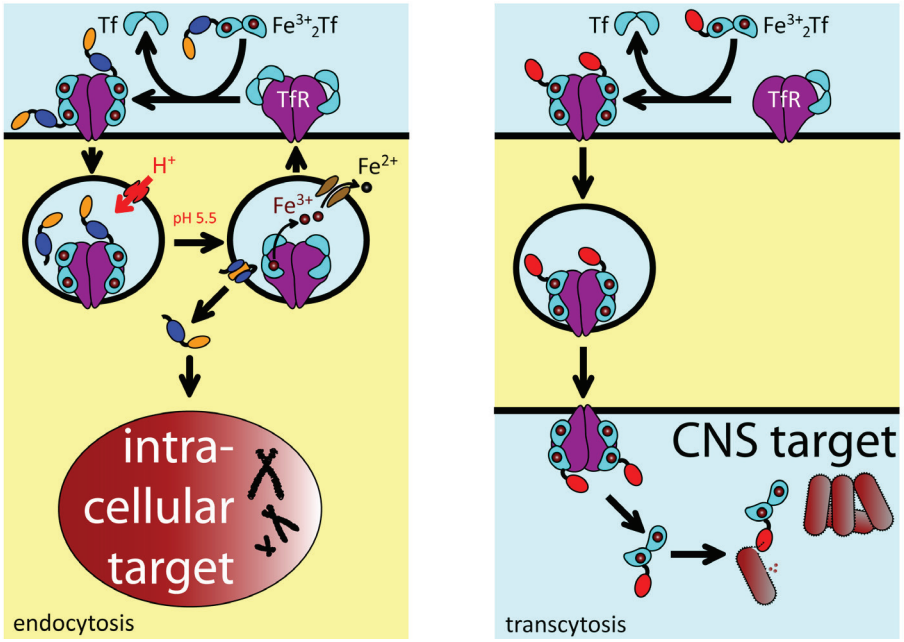


Figure 1-1. Mechanism of targeting tumor cells (left) and central nervous system (right) using TfR-mediated endocytosis/transcytosis. Figure from Kaltashov, I.A., et al., Adv Drug Deliver Rev, 2013, accepted.

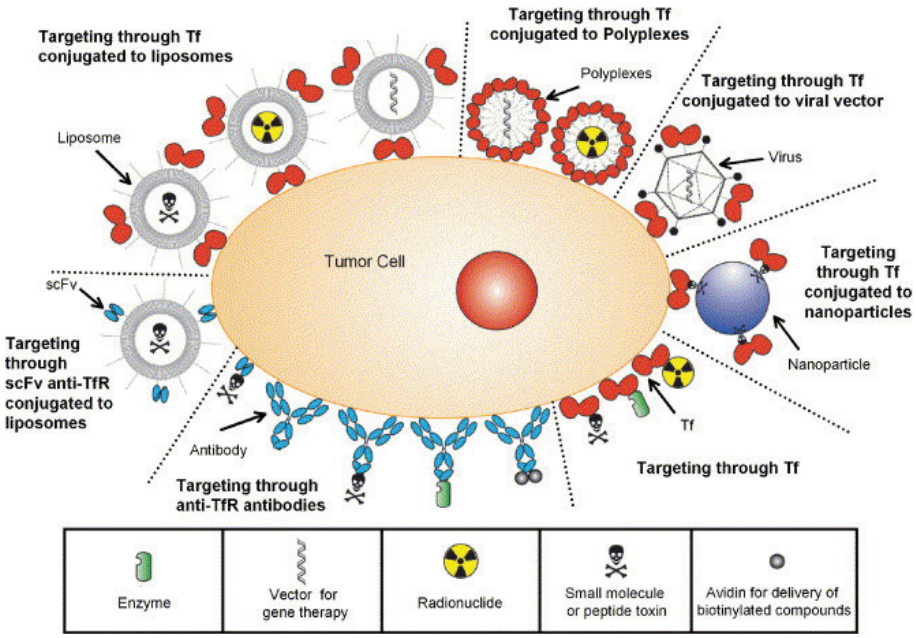


Figure 1-2. Tf-based strategies to deliver various therapeutic agents to tumor cells via TfR-mediated endocytosis. Figure is from reference [11].

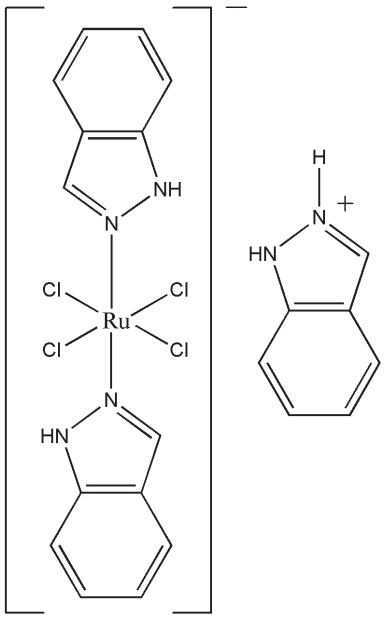


Figure 1-3. Structure of anticancer agent KP1019.

CHAPTER 2

DEVELOPMENT OF MS-BASED METHOD FOR QUANTITATION OF HUMAN TRANSFERRIN IN BIOLOGICAL FLUIDS

2.1 Introduction

Rapid growth of quantitative proteomics in which mass spectrometry (MS) has played a central role provides a new strategy for protein quantitation in a both sensitive and accurate fashion. By employing differential stable isotope labeling of proteins or peptides in two sample sets, mass differences can be created and recognized by MS analysis. Absolute quantitation of protein can be achieved by comparing the MS signal from the biological/clinical sample with the signal from the spiked internal standard present at a known concentration (Figure 2-1). More recently, a label-free quantitation method by MS has emerged as an alternative approach [31], and gained wide applications in quantitative proteomics. However, the most accurate quantitation is normally achieved using stable isotope labeled internal standards.

2.1.1 Stable isotope labeled internal standard

For accurate quantitation, it is essential to use an isotopically labeled internal standard to compensate for the matrix effects during LC-MS analysis. Based on when and how the isotopes are introduced, the internal standards can be divided into two categories: protein-based internal standard and peptide-based internal standard.

Apparently, for most accurate quantitation, protein-based internal standard is more ideal than peptide-based internal standard, since the former can be spiked into biological samples at a very early stage. As a result, all the downstream sources of quantitation error can be effectively removed (Figure 2-2) [32].

For example, isotopically labeled protein can be produced with relative ease in prokaryotic expression system during cell culture using ^{15}N-enriched medium, which is the earliest possible point for introducing isotopes [33]. However, production of labeled protein in eukaryotic expression system by this approach is significantly more costly and labor-intensive. Instead, stable isotope labeling by amino acid in cell culture (SILAC) approach introduced by Mann and co-workers in 2002 has gained wider popularity [34]. Most frequently, $^{13}C_6$-arginine and $^{13}C_6$-lysine are used in the medium so that all the tryptic fragments of a protein carry at least one labeled amino acid, resulting in a constant mass increment that can be analyzed by MS. Alternatively, protein-based internal standard can also be produced by chemical labeling at the intact protein level. Amine-based labeling strategies (Lys, N-termini) are less favorable due to the frequent application of trypsin during protein proteolysis. As a result, thiol-based labeling strategies are the primary way to introduce isotopes at the intact protein level. For example, the ICAT approach has been widely applied for quantitative analysis of complex protein mixtures, where the thiol groups in protein were derivatized with isotope-coded reagents [35].

In comparison with isotopically labeled protein which is significantly more difficult

to produce, isotopically labeled peptides can be produced with relative ease via a variety of approaches. Indeed, the application of peptide-based internal standard in protein quantitation is much more frequent than the use of protein-based internal standard. Conveniently, isotopically labeled peptide can be synthesized based on a reference peptide from proteolysis of the protein of interest. Thus, absolute quantitation of the protein can be achieved by comparing the signal intensity from unlabeled (biological sample) and labeled peptide (internal standard). This approach was first introduced by Gygi and co-workers in 2003 termed AQUA [36], and it has been the most widely applied technique for protein quantitation. In addition, isotopes can also be introduced to peptides by chemical labeling. Particularly, amine groups (Lys, N-termini) are primarily targeted and can be derivatized with various labeling reagents, including iTRAQ [37], dimethyl labels [38] and tandem mass tag (TMT) reagents [39]. All those approaches have gained wide acceptance in the field of quantitative proteomics. Alternatively, enzyme-catalyzed O^{18} labeling is another elegant way to introduce isotope labeling to proteolytic fragments. Yao and co-workers [40] first applied this technique in a comparative proteomics study. Despite the ease of use, application of this approach for accurate protein quantitation is hampered by the back exchange event during LC-MS analysis due to the residual protease activity in the system, although a lot of efforts have been devoted to reduce this effect [41-44]. In addition, an intrinsic disadvantage of enzyme-catalyzed O^{18} labeling is the peptide-dependant O^{18}-incorporation efficiency, which further complicates the data analysis [45]. More recently, a modified approach

termed acid-catalyzed O^{18} labeling was introduced, which takes advantage of free exchange of carboxyl oxygens in peptides with solvent at low pH [46]. Despite the prolonged preparation time, this approach provides sufficient mass shift after labeling due to additional O^{18}-incorporation sites on the acidic residues. Furthermore, since no enzyme is employed during the labeling, back exchange rates could be significantly reduced. However, despite this variety of techniques, peptide-based internal standard is not ideal for protein quantitation, since it can only be spiked into biological samples after proteolysis, which inevitably introduces sources of quantitation error. In particular, digestion efficiency becomes a significant variable that can greatly affect the quantitation results. In fact, complete digestion of target protein is normally required for reliable quantitation, although it is frequently difficult to achieve.

Isotope effect is another important factor that needs to be taken into consideration during the selection of isotopes. Although it is generally valid that isotope labeling does not alter the physicochemical properties of a peptide, retention time shifts during RP-HPLC separation have been reported for deuterated peptides [47]. Obviously, identical retention time between unlabeled and isotope-labeled peptides is critically important for accurate quantitation, to ensure the identical solvent composition and ion suppression from co-eluting species, resulting in identical ionization efficiency of unlabeled and labeled peptides. Thus, instead of deuterium, isotope labels such as C^{13}, N^{15}, or O^{15} have been more widely applied, due to their negligible isotope effect.

2.1.2 Enrichment of human Tf

Since a protein drug in biological fluids typically represents a very small fraction of the total proteome, its detection by MS or LC-MS can be hampered by both insufficient signal intensity (ion suppression) and interference caused by co-eluting components of similar mass. The most straightforward way to overcome this difficulty is to reduce sample complexity by either removing the most abundant endogenous proteins or enriching the target protein *via* affinity-based purifications. For example, up to 21 high abundant proteins can be removed from plasma sample with relative ease using commercially available immunodepletion columns [48, 49]. Alternatively, if an appropriate antibody is available, the protein analyte can be simply enriched in one-step applying immunocapture/immunoprecipitation [50]. However, both those methods are less useful for enrichment of exogenous Tf (e.g., human Tf) from biological fluids due to cross-reactivity from abundant endogenous Tf (e.g., bovine Tf or rat Tf) or *vice versa*.

Instead, immobilized metal ion affinity chromatography [51] is employed in this work, which takes advantage of the coordinate binding of amino acids, particularly histidine, to metal ions, such as cobalt, nickel and copper [52]. Particularly, nitrilotriacetic acid (NTA) is applied as chelating ligand to retain the metal ions on the solid surface. Due to the strong binding ability of NTA to metal ions, a wide variety of binding/washing conditions (particularly stringent washing conditions) can be employed in order to enrich low abundance target protein from complex matrices. After

introducing a 6×His-tag to the N-terminus of hTf (referred to as Tf*), enrichment of Tf* can be achieved by applying Ni-NTA affinity purification.

2.1.3 LC-MS based protein quantitation

LC-MS based method using conventional separation conditions (4.6-mm i.d. column and 1.0 mL/min flow rate or 2.1-mm i.d. column and 0.2 mL/min flow rate) has been a mainstay for quantitation of small molecular compounds from biological fluids due to the high-throughput and robustness. However, for detection and quantitation of low levels of therapeutic proteins in complex matrices, LC-MS analysis with standard flow rate separation normally cannot meet the requirement of high sensitivity. Instead, capillary flow (1-10 µL/min) and nano flow (ca. 200 nL/min) are frequently applied in order to enhance the sensitivity by reducing dimensions of the system [53, 54]. In our study, a nanoLC-MS system interfaced by a nano ESI spray source is applied to provide necessary sensitivity for the quantitation of exogenous Tf* in biological fluids. To cope with the nano flow, digestion samples (ca. 50-200 µL) are first preconcentrated using a micro-trapping column at a relatively high flow rate (40 µL/min), and subsequently eluted to the analytical column at a low flow rate (200 nL/min). A control experiment showed that amount as low as 8 ng of Tf* can be confidently detected using this setup.

2.2 Results and discussion

2.2.1 MS-based method for quantitation of Tf* in bovine serum

In a model study, method consisting of Ni-affinity purification for protein enrichment, acid-catalyzed O^{18} labeling for preparation of internal standard and nanoLC-MS detection was developed for quantitation of low levels of Tf* in bovine serum.

2.2.1.1 Ni-NTA affinity purification

In order to improve the sensitivity of the method, Ni-NTA affinity purification was employed to extract low abundance Tf* from bovine serum samples. Briefly, 50 µL of Ni-NTA resin slurry was washed by 1 mL of binding buffer, and then transferred to the sample solution in a 1.5 mL Eppendorf tube. The binding was conducted at room temperature for 90 min with gentle shaking. Subsequently, the resin was washed by 500 µL of washing buffer (50 mM of potassium phosphate, 20 mM of imidazole, 6 M of guanidine hydrochloride, 1 M of sodium chloride and 10% v/v glycerol, pH 8.0) for two times in order to remove non-specific bound proteins. Finally, Tf* was recovered by adding 100 µL of elution buffer (50 mM of potassium phosphate, 300 mM of imidazole, pH 8.0) to the resin, and subsequently collecting the supernatant. The recovery of Tf* after Ni-affinity purification was investigated using size exclusion chromatography (SEC), and calculated to be ca. 80% based on the UV signal of Tf* peak (Figure 2-3).

After Ni-affinity purification, the complexity of the serum sample was significantly reduced, which can be clearly observed in SDS-PAGE analysis (Figure 2-4). By comparing the protein bands from the flow-through fraction and the purification fraction, it was clear that majority of serum proteins were efficiently removed after Ni-affinity purification.

The enrichment of Tf* from the bovine serum sample by Ni-affinity purification was further studied using a "bottom-up" approach. Briefly, bovine serum samples spiked with Tf* (1 μg of Tf* in 1 mL of bovine serum) were treated with or without Ni-affinity purification. After tryptic digestion, proteolytic peptides were subjected to LC-MS/MS analysis, and the results were searched against SwissProt protein database by Mascot MS/MS ions search engine [55] to identify the precursor protein(s). The top five protein hits in each sample were listed in Table 1. Without Ni-affinity purification, Tf* cannot be successfully identified due to the presence of abundant endogenous proteins (complete list was not shown in Table 1). In contrast, with the purification step Tf* can be confidently identified, emphasizing the necessity of applying Ni-affinity purification for detection of low levels of Tf* in bovine serum samples.

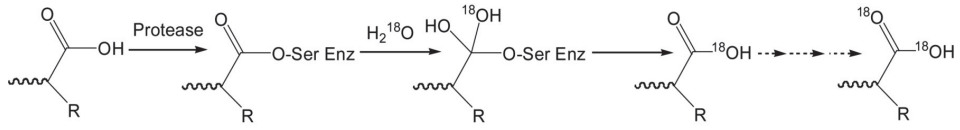
2.2.1.2 Preparation of internal standard by acid-catalyzed O^{18} labeling

O^{18}-labeling was selected for preparation of stable isotope labeled internal standard in our study due to the minimal isotope effect and the ease of use. Two classic approaches to introduce O^{18}-labeling have been reported, termed "enzyme-catalyzed

O^{18} labeling" (Scheme 1) and "acid-catalyzed O^{18} labeling" respectively (Scheme 2). Enzyme-catalyzed O^{18}-labeling is widely applied in quantitative proteomics. It takes advantage of the reversible binding between enzyme (mostly trypsin) and the proteolytic peptides in H_2O^{18} so that both the C-terminal carboxylic oxygens can be exchanged to O^{18}, leading to a mass increment of 4 Da for each peptide. Despite the ease of use, this approach is limited by some intrinsic problems. First, complete labeling is difficult to achieve, resulting in overlap of isotopic peaks with unlabeled peptides. In addition, peptide-dependant labeling efficiency greatly complicates the data analysis [45]. Furthermore, back exchange of O^{18}-labeled peptides occurring in RP-HPLC separation can be significant due to the catalysis of residual trypsin. As a result, removal or inactivation of trypsin after labeling is normally required. Most of those problems can be effectively eliminated by increasing the mass shift after labeling. Indeed, the new approach, acid-catalyzed O^{18} labeling, explored the possibility of labeling all the carboxylic acid groups (C-terminus, acidic residues) in order to increase the mass shift [46]. By incubating the tryptic fragments in H_2O^{18} under acidic condition, all the carboxyl oxygens can be exchanged to O^{18} upon equilibrium. Owing to the high frequency of acidic residues in the human proteome, nearly half of the tryptic fragments contain at least one acidic residue, providing a mass shift of 8 Da or more, which is sufficient to completely separate isotopic clusters from unlabeled and labeled peptides [46]. In addition, as no enzyme is required during the labeling, the back exchange rate of labeled peptides is also significantly reduced. However, this approach requires a considerably

longer preparation time of the internal standard (typically over 10 days), thus careful planning of the experiments is generally suggested. Both approaches were investigated in our study.

Scheme 1



Scheme 2

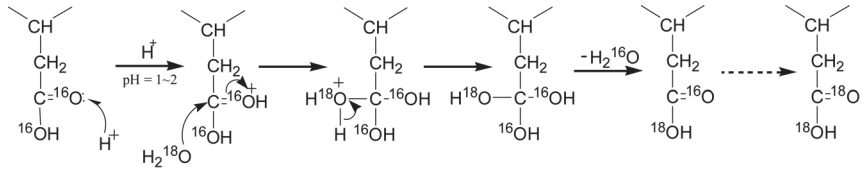


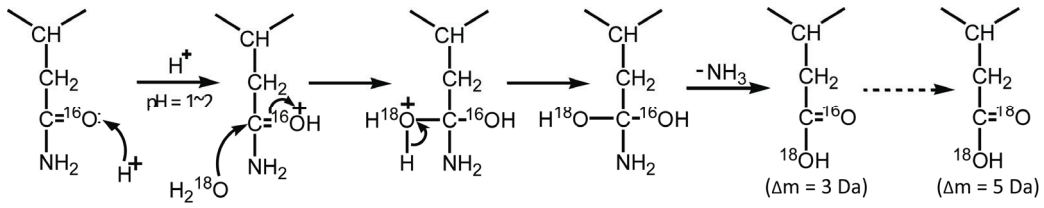
Figure 2-5 represents the labeling results of a tryptic fragment from Tf* by both trypsin-catalyzed O^{18} labeling (blue trace) and acid-catalyzed O^{18} labeling (red trace), in comparison with unlabeled peptide (black trace). Clearly, sufficient mass increment provided by acid-catalyzed O^{18}-labeling greatly simplifies the data analysis. Thus, this approach was selected to prepare an O^{18}-labeled internal standard for quantitation of Tf* in bovine serum.

2.2.1.3 Discovery of pitfalls in protein quantitation using acid-catalyzed O^{18} labeling

The quantitation method for Tf* in bovine serum was developed using acid-catalyzed O^{18} labeling. However, pitfalls of this approach were identified during the initial method development [56]. As shown in Figure 2-6, quantitation of Tf* in bovine serum was achieved using the identical amount of O^{18}-labeled internal standard, and

results were generated using several reference peptides (black bars). Surprisingly, a large discrepancy in quantitation results was observed among different reference peptides. Furthermore, even though the unlabeled to labeled ratio should not exceed 1.0 (preparation of internal standard does not expose Tf* to the affinity purification step, thereby eliminating unavoidable protein loss), the apparent ratio for one of the peptides shown in Figure 2-6 exceeded that level. Close examination of the data presented in Figure 2-6 revealed a common structural feature among all peptides that yielded higher quantitation levels of Tf * in bovine serum: each of these peptides has at least one amino acid residue bearing an amide group on its side chain (Asn, Gln, and carboxamidomethylated cysteine, Cys*). Because of the susceptibility of amides to acid hydrolysis, acid-driven deamidation was suspected as a likely cause of the discrepancy in quantitation results (Scheme 3). Indeed, if the fraction of deamidated peptides (from O^{18}-labeled internal standard) was neglected during the ratio calculation, the apparent level of Tf* in bovine serum would necessarily be inflated. In addition, acid-driven deamidation in H_2O^{18} should add another 5 Da to the labeled peptides due to additional incorporation of two O^{18} atoms (Scheme 3).

Scheme 3



An example of this phenomenon is shown in Figure 2-7. For peptide EFQLFSSPHGK,

a maximum mass increment of 8 Da was expected after acid-catalyzed O^{18} labeling, due to O^{18} incorporation at both the glutamic acid residue and the C-terminus. Simultaneously, acid-driven deamidation at glutamine residue led to additional incorporation of two O^{18} atoms, resulting in a total mass increment of 13 Da. LC-MS analysis revealed the consistent results. In addition to unlabeled peptide (I_0) and O^{18}-labeled peptide (I_8) co-eluted at the first peak, another ionic species eluted slightly later with maximum mass increment of 13 Da (I_{13}) was also observed.

The deamidation site was further confirmed by tandem MS experiment (Figure 2-8). Collisional activation of the doubly charged ion corresponding to this peptide yielded a series of single charged y-ions. The fragment ion mass spectrum of the I_{13} species did not show any additional mass shifts compared to that of I_8 for smaller fragments (up to y_8), suggesting that no modification was present in the last eight residues. However, an additional mass shift of 5 Da was observed for fragment ion y_9, clearly suggesting that deamidation of Glu followed by incorporation of two O^{18} atoms into the newly formed carboxylic acid group. Similar analyses were carried out for other amide-containing peptides, and in all cases extensive deamidation was observed in internal standard prepared by acid-catalyzed O^{18}-labeling (data not shown). The extent of deamidation varied among different peptides, but on average it was within 15-30% of the total peptide population following 2 weeks of acid-catalyzed labeling (the typical amount of time needed to achieve a near-complete labeling of the acidic side chains).

Clearly, the loss of a significant fraction of an internal standard due to deamidation

led to an artificial increase of the apparent level of Tf* in bovine serum. An attempt to remediate this problem by including all deamidated species in the ratio calculation led to a significant decrease in the discrepancy of the quantitation results (Figure 2-9), but did not eliminate it completely. One important reason for the remaining discrepancy is the inadvertent change of the ionization efficiency of deamidated peptides compared to their intact counterparts (which is caused not only by the alteration of their covalent structure but also by a change in the solvent composition at different retention time and a different matrix/ion suppression effect from coeluting species). Another reason is the existence of a matrix effect during the tryptic digestion of serum sample and the internal standard, since the latter was prepared in the absence of bovine serum. However, this matrix effect on digestion efficiency is a universal problem for nearly all the quantitation methods using peptide-based internal standard, which can only be eliminated by switching to protein-based internal standard.

2.2.1.4 Quantitation of Tf* in bovine serum

As demonstrated in this study, selection of reference peptides is critically important in this approach. Apparently, amide-containing peptides should be excluded due to the susceptibility to acid-driven deamidation. In addition, for methods using peptide-based internal standard, complete digestion is required to minimize the matrix effect on digestion efficiency. Thus, only peptides without any missing cleavage can be used as reference peptides. And finally, the use of high abundance peptides can effectively

improve method sensitivity. Using two such peptides, calibration curves were generated for quantitation of Tf* in bovine serum (Figure 2-10). Excellent linearity was achieved for over two orders of magnitude in each case. Concentrations as low as 0.2 μg/mL of Tf* can be confidently quantitated using this approach, which is ca. 4,000 times lower than the concentration of endogenous bovine Tf.

2.2.2 Cysteine O^{18}-labeling for accurate protein quantitation

As discussed earlier and demonstrated in our first study, the application of peptide-based internal standard is not ideal for accurate protein quantitation. Particularly, the matrix effect on digestion efficiency can greatly compromise the quantitation results, since the digestion of biological sample and internal standard are performed separately. A logical solution for this problem is the use of a protein-based internal standard so that the addition of the internal standard to the biological sample can occur prior to digestion. However, classic approaches (enzyme-catalyzed O^{18}-labeling and acid-catalyzed O^{18}-labeling) only introduce the isotopes at the peptide level, thus are both suffering from this disadvantage. Hereby, we developed a new strategy of O^{18}-labeling at the intact protein level, termed cysteine O^{18}-labeling, in which the isotope incorporation occurs during the alkylation of cysteine residues. The feasibility of the new method for accurate protein quantitation has been demonstrated using Tf* as a model.

2.2.2.1 The workflow of cystein O^{18}-labeling for protein quantitation

The workflow of using cysteine O^{18}-labeling for protein quantitation is represented in Figure 2-11. Briefly, after reduction of disulfide binds in both biological samples and internal standard, the cysteine residues were alkylated with either O^{16}-labeled or O^{18}-labeled iodoacetic acid. Subsequently, an internal standard of known quantity was spiked into the biological samples, followed by tryptic digestion of the mixture. Finally, the protein quantitation was achieved by comparing signal intensities of differentially alkylated products from a cysteine-containing peptide.

2.2.2.2 Preparation of O^{18}-labeled iodoacetic acid (IAA)

The O^{18}-labeled IAA was prepared by exchanging its oxygen atoms in carboxylic group with H_2O^{18} at low pH. The labeling results are shown in Figure 2-12, where the isotopic distribution of IAA after a 24 hour incubation at 50 °C, in the presence of 1 % TFA, shows > 85 % exchange of both carboxylic oxygen atoms and > 98 % exchange of at least one oxygen atom. These numbers yield the overall extent of O^{18} incorporation as 93 %, whereas the theoretical value is calculated to be 93.9% after taking into account the presence of residual O^{16} in the reaction mixture. This consistent result suggested that the O^{18} exchange between IAA and water had reached the equilibrium.

2.2.2.3 Cysteine O^{18}-labeling: a model protein study

The alkylation of cysteine residues with O^{18}-labeled IAA was first studied using Tf* which has 38 cysteine residues as a model protein. Figure 2-13 presents an LC-MS/MS analysis of the mixture of tryptic fragments from two batches of Tf* capped by

O^{16}-labeled and O^{18}-labeled IAA, respectively. The mass spectrum (Figure 2-13B) of a tryptic peptide KPVEEYANC*HLAR indicates a maximum mass shift of 4 Da between unlabeled (I_0) and labeled (I_4) peptides. Extracted ion chromatograms plotted for their most abundant isotopic peaks, I_0 and I_4 (Figure 2-13A), clearly show identical elution times, which is critically important for reliable quantitation. The tandem MS data (Figure 2-13C and D) suggest that the presence of IAA moiety on a peptide ion does not interfere with its identification using CAD (contrary to ICAT labeling [57]). Furthermore, the isotopic distributions of y^{4+} and y^{5+} ions provide a clear indication as to the incorporation site of O^{18} atoms within the peptide (cysteine residue).

2.2.2.4 Labeling efficiency of cysteine O^{18}-labeling

In this approach, O^{18} labeling is introduced by alkylation reagent, thus the labeling efficiency of all the cysteine-containing tryptic fragments should be consistent. This assumption was demonstrated by calculating the labeling efficiency of all the detectable cysteine-containing peptides from tryptic digestion of Tf* (capped by O^{18}-labeled IAA). As shown in Table 2, for peptides containing one cysteine residue, the full labeling products (with both two O^{18} atoms incorporated) made up to 87%–90%, consistent with the labeling results of IAA (*vide supra*). Unlike enzyme-catalyzed O^{18} labeling which has peptide-specific O^{18} incorporation rate, the use of O^{18}-labeled IAA generated almost identical labeling efficiency among all the cysteine-containing peptides. This feature greatly simplifies the data analysis, particularly in the case of insufficient mass shift

(incomplete separation of isotopic distributions of labeled and unlabeled peptides). In fact, the ratio between unlabeled and labeled peptides can be easily estimated by the following equation:

$$peptide\ \ ratio = \frac{I_1}{I_5 \times (1+\alpha)}$$

Where α (α=0.14 in this work) is a fraction of partially labeled peptides (only one O^{18} atom incorporated); and I_1 and I_5 represent the observed relative intensities of the second isotopic peaks for the unlabeled peptides and fully O^{18}-labeled peptides, respectively.

2.2.2.5 Stability of cysteine O^{18}-labeling

The back exchange in trypsin-catalyzed O^{18} labeling is primarily contributed from the residual trypsin activity in the labeling solution. However, in this new approach, the O^{18}-incorporation sites (carboxymethylated Cys) cannot be recognized by trypsin, thus the back exchange rate is expected to be extremely low (without the catalysis of trypsin). To validate this assumption, stability of the cysteine O^{18}-labeled peptide KPVEEYANC*HLAR was investigated at both RP-HPLC conditions (0.1% FA) and near neutral pH (pH 8, for tryptic digestion) (Figure 2-14). The relative abundance of the isotopic peak I_2 as percentage of the total ionic signal was calculated and monitored over time. No noticeable back exchange was observed under RP-HPLC conditions for at least 8 hours, or under near neutral pH for at least 6 days. The excellent stability of O^{18}-labeling under near neutral pH allows overnight tryptic digestion and any other

sample processing step (as long as it does not require a long period of time in extreme pH conditions) after spiking the internal standard. In addition, the good stability of O^{18}-labeling under RP-HPLC conditions allows us to apply a long LC gradient separation for complex samples.

2.2.2.6 The feasibility of using cysteine O^{18}-labeling for absolute protein quantitation

To demonstrate the feasibility of using cysteine O^{18}-labeling for protein quantitation, we generated Tf* samples capped with unlabeled and O^{18}-labeled IAA at various ratios covering three orders of magnitude, and performed nano-LC/MS analysis. Successful quantitation was achieved using two cysteine-containing tryptic peptides from Tf* respectively (Figure 2-15). High correlation of theoretical and experimentally observed ratios was achieved ($R^2 > 0.99$). Two reference peptides both showed good linearity in up to three orders of magnitude. Furthermore, the quantitation results generated by these two peptides were highly consistent, highlighting the reliability of quantitation using protein-based internal standard.

2.2.2.6 Conclusion

Obviously this new method only targets cysteine-containing tryptic peptides and therefore it cannot be applied to quantitate proteins without cysteine residue. Nevertheless, even the most conservative estimates of the fraction of completely cysteine-free proteins in the proteomes of higher organisms put this number as low as 8%, while the SWISS-Prot database puts the number of human proteins that have at

least one cysteine residue in the sequence as high as 97 %. As a result, this method can still be widely applied for tasks ranging from pharmacokinetics studies to quality control of biopharmaceutical products, where accurate protein quantitation is required.

2.2.3 Development of MS-based method for quantitation of Tf* in rat CSF

In spite of the success achieved by protein drugs in treating human diseases in the last decade, development of protein drugs targeting CNS has been notably slow, primarily due to the presence of blood-brain barrier (BBB). As one of the very few proteins that can penetrate the BBB via receptor-mediated transcytosis, Tf has been a focus of extensive efforts to achieve delivery of therapeutic agents to CNS [58]. In order to investigate the BBB-traversing ability of Tf, quantitation method for low levels of exogenous Tf in CSF samples is required. As discussed in section 1.1.5, immunoaffinity-based assays are less useful for this task due to the presence of abundant endogenous Tf in CSF. Although radioactive labeling based approaches provide extremely high sensitivity, protein quantitation is achieved indirectly by measuring the radioactivity, and therefore might be less accurate and reliable. Hereby, we developed an LC-MS based method for accurate and sensitive quantitation of Tf* in rat CSF, where the newly developed cysteine O^{18}-labeling approach was applied to produce internal standard.

2.2.3.1 Ni-NTA affinity purification of Tf* from rat CSF

Although CSF presents a significantly less complex system compared to plasma

(contains approximately 0.3% plasma proteins [59]), an enrichment step for exogenous Tf* is still required. Indeed, a similar "bottom-up" analysis (as discussed in section 1.2.1.1) of 1 μg of Tf* in 50 μL of rat CSF sample demonstrated the necessity of applying Ni-NTA affinity purification prior to LC-MS detection (data not shown). In addition, a comparative analysis of the flow-through fraction and purification fraction from Ni-affinity purification was also conducted to investigate the enrichment of Tf* and removal of endogenous proteins. Briefly, control sample with 1 μg of Tf* in 20 μL of rat CSF was treated with Ni-affinity purification. Cysteine residues from proteins in flow-through fraction and purification fraction were capped by O^{16}-labeled iodoacetic acid (IAA-16) and O^{18}-labeled iodoacetic acid (IAA-18) respectively. Subsequently, after combination of the two fractions, the sample was digested by trypsin, and subjected to LC-MS/MS analysis. By comparing the relative intensities of differentially alkylated peptides (ion I_0 from flow-through fraction and ion I_4 from purification fraction) from certain protein, the enrichment or removal of this protein can be evaluated. Clearly, after Ni-affinity purification, Tf* was greatly enriched to the elution fraction; in contrast, two of the most abundant endogenous proteins in CSF, rat Tf and rat albumin, were effectively removed (Figure 2-16).

2.2.3.2 Quantitation of Tf* using cysteine-containing peptides

O^{18}-labeled Tf* was used as internal standard for quantitation, which was prepared by alkylation of cysteine residues with IAA-18. Thus, only cysteine-containing peptides

could be used as reference peptides. Figure 2-18 represents an example of using a tryptic peptide with two cysteine residues for protein quantitation. Since the isotopic distributions from unlabeled and labeled peptides are completely separated due to sufficient mass shift (with a maximum of 8 Da), an m/z range (the grey area shown in Figure 2-17) covering all the isotopic peaks from either unlabeled or labeled peptide is extracted. Subsequently, quantitation can be achieved using a simple equation (Figure 2-17) with peak areas derived from extracted ion chromatograms (XICs).

In another case, protein quantitation can also be achieved using a reference peptide containing a single cysteine residue (Figure 2-18). Although isotopic overlapping between unlabeled and labeled peptides can be observed due to incomplete labeling, the interference can be simply corrected by introducing a constant (α) that represents the fraction of partially labeled peptide (with only one O^{18} atom incorporated). To further minimize the interference, the second isotopic peak of either unlabeled peptide (I_1) or fully O^{18}-labeled peptide (I_5) are extracted, and the peak areas under XICs are used for the ratio calculation (Figure 2-18).

2.2.3.3 Quantitation results by different reference peptides

A great advantage of using protein-based internal standard is to eliminate the discrepancies in digestion efficiency between biological sample and internal standard. This assumption was demonstrated by comparing quantitation results generated by different peptides, including those with missing cleavages. Triplicate samples containing

1.6 µg of Tf* in 50 µL of rat CSF were spiked with 0.4 µg of internal standard, and analyzed by this method. Ten cysteine-containing peptides were used for quantitation, and the results were normalized against the average value shown in percentage (Figure 2-19). Clearly, nearly all the quantitation results were well within ±15% range regardless of the peptide used. Importantly, even peptides with missed cleavages showed highly consistent quantitation results. Those observations indicated that the discrepancies in digestion efficiency were effectively reduced by using the protein-based internal standard. Furthermore, the feasibility of using peptides with missing cleavages for protein quantitation greatly expanded the selection of reference peptides. In fact, peptide DLLFRDDTVC*LAK was selected as a reference peptide in this study since it was unique to Tf* (in comparison with rat Tf) and could be consistently detected at low concentration, although it did contain a missed cleavage.

2.2.3.4 Calibration curve and method validation

Using peptide DLLFRDDTVC*LAK as the reference peptide, a linear calibration curve was plotted using measured concentration of Tf* versus the corresponding theoretical spiked concentration (20, 40, 100, 200, 800, 1600, and 3200 ng of Tf* in 50 µL of rat CSF). A $1/X^2$ weighted regression was applied. Triplicate samples were measured for each data point, and the standard deviation was presented as the error bar. As shown in Figure 2-20, excellent linearity was achieved for more than two orders of magnitude.

Accuracy and precision. The accuracy and precision of this method were further

validated using QC samples prepared with concentrations in the low, medium and high parts of the calibration curve. As shown in Table 3, both precision (% CV) and accuracy at all quality control levels were well within 15%, indicating that the method is both precise and accurate.

Selectivity and lower limit of quantitation (LLOQ). The uniqueness of the reference peptide DLLFRDDTVC*LAK for Tf* was confirmed by a protein database search against UniProtKB/Swiss-Prot using the Protein BLAST search engine [60]. In addition, blank samples containing 50 µL of rat CSF were analyzed in triplicates by this method. No signal was observed at the retention time of the reference peptide. During the method development, the lower limit of quantitation of this method was estimated to be 20 ng of Tf* in 50 µL of rat CSF. Measurement of QC samples with this concentration demonstrated that both good precision (% CV) and accuracy (within 20%) can be achieved at LLOQ level (Table 3).

Matrix effect. It is well known that the total protein level in rat CSF can vary significantly from individual to individual [61]. Thus, in order to apply this method in animal studies, it is critically important to investigate the matrix effect. QC samples containing 500 ng of Tf* in 100 µL of rat CSF (the calibration standards were prepared using 50 µL of rat CSF) were analyzed in triplicates, and the results were compared with the value derived from the calibration curve. As shown in Table 3, both accuracy and precision of the measurements were well within 15% range, suggesting negligible matrix effect within two times of complexity.

1.2.3.5 Conclusion

Applying cysteine O^{18}-labeling, an LC-MS based method was developed and validated for quantitation of Tf* in rat CSF. The newly developed labeling strategy allows the preparation of a protein-based internal standard, which proved to be superior for MS-based quantitation. Excellent linearity, accuracy and precision were achieved using this method. In addition, minimal matrix effect was found after increasing the sample complexity by two times, suggesting the feasibility of applying this method in real animal studies. Finally, concentrations as low as 10 ng of Tf* in 50 µL of rat CSF can be confidently quantitated, providing necessary sensitivity for quantitation of low levels of Tf* in rat CSF after intravenous injection.

2.2.4 Indium labeling and ICP-MS detection based method for Tf quantitation

In addition to LC-MS based approach, we also explored the possibility of using inductively coupled plasma mass spectrometry (ICP-MS) for quantitation of Tf* in biological fluids. Using indium as a tracer, selective and sensitive quantitation of Tf* can be achieved by detection of indium by ICP-MS. This work was done in collaboration with Hanwei Zhao in our lab. Hanwei prepared the Tf bound with different metal ions, and investigated their stability and binding ability to TfR. I then acquired the ICP-MS spectra of In_2Tf in serum samples, and developed and validated the quantitation method for indium in serum and rat CSF.

2.2.4.1 Metal labeling and ICP-MS detection based methods for protein quantitation

ICP-MS is an analytical tool widely used for elemental analysis, which consists of a high-temperature plasma for sample ionization and MS for ion separation and detection. During the last decade, ICP-MS has greatly expanded its applications into a wide variety of areas, particularly in the field of bioanalysis [62]. For example, it was reported that ICP-MS based methods have been applied for protein quantitation using a metal labeling strategy. Frequently, labeling of the protein with rare-earth metals is preferred, which allows sensitive quantitation due to their high ionization efficiency and extremely low background (absence of rare-earth elements in biological samples and no spectroscopic interferences). Figure 2-21 shows the structure of two metal chelates that are most frequently employed for protein derivatization. Both chelates DTPA [63] and DOTA [64] can form strong complexes with multiple lanthanides (e.g., log K 22.5 for Gd[DTPA] and log K 24.7 for Gd[DOTA] [65]), providing a possibility to achieve multiplexing. However, protein labeling efficiency with those metal chelates is generally low due to their relatively large size. In addition, extensive modification of a protein by metal chelates can significantly alter its biological activity or function, and therefore such studies have not explored the possibility of using metal-coded protein drugs for pharmacokinetic studies.

2.2.4.2 Indium as a tracer for Tf* quantitation

Instead of using metal chelates, we took advantage of the two metal binding sites in Tf. It is well known that Tf can bind to several other metal ions in addition to iron,

including aluminum, gallium, indium, and thallium [25]. The use of ^{67}Ga- and ^{111}In-bound Tf has even gained widespread medical and radiodiagnostic applications [26-28]. In our study, indium was selected as the tracer due to its high binding affinity to Tf. High resolution MS analysis by Hanwei Zhao in our lab could confidently discriminate apo-Tf and Tf bound with different metal ions based on accurate masses. MS analysis confirmed that indium binds to Tf with a constant ratio of 2:1, which provides an easy conversion between indium concentration and Tf concentration. In addition, indium-bound Tf was observed to be stable in the presence of ferric ions or other proteins (e.g. albumin), demonstrating the feasibility of using indium as a tracer for Tf in the circulation (data now shown). Furthermore, our study also revealed that indium-bound Tf had similar binding affinity to Tf receptor (low μM range) as diferric Tf, which is critically important to perform its biological functions (data not shown). At last, the absence of indium in all living systems provides an extremely low background during ICP-MS analysis, and therefore greatly improves the method sensitivity.

2.2.4.3 ICP-MS analysis of serum sample spiked with In_2Tf

Figure 2-22 shows a full-scan spectrum of serum sample spiked with In_2Tf acquired by ICP-MS. Signal from elements within 50 to 150 m/z range was recorded and shown in black trace. Colored bars represented the natural isotopic distributions of several elements. The excellent agreement between theoretical distribution and measured distribution further confirmed the identity of each element. As shown in the inset, low level of indium (0.5 ppb) can be readily detected without interferences from other

elements, suggesting good selectivity of this method.

2.2.4.4 Quantitation methods for indium in serum and CSF

Applying the new strategy, quantitation of Tf was converted to quantitation of indium in biological fluids, which can be easily achieved with high sensitivity even without purification step. Another great feature of ICP-MS analysis is the minimal matrix effect due to the complete ionization induced by high plasma temperature. In this study, 2.5% of nitric acid solution was used as surrogate matrix of serum and CSF to generate calibration curves (Figure 2-23). In both cases, excellent linearity was achieved for nearly three orders of magnitude.

Subsequently, the quantitation methods were validated for accuracy, precision and matrix effect. And all of them were well within 15% range (Table 4). Concentration as low as 0.02 ppb of indium can be confidently quantitated in CSF samples, providing necessary sensitivity for quantitation of low levels of In_2Tf in rat CSF after intravenous injection.

2.3 Conclusion

LC-MS based methods were developed and validated for quantitation of low levels of Tf* in biological fluids. Particularly, pitfalls in protein quantitation using a classic approach (acid-catalyzed O^{18} labeling) were discovered, and successful quantitation was achieved after circumventing those problems. In addition, a new strategy of O^{18}-labeling was developed to produce protein-based internal standard, which greatly improved the reliability of the quantitation results. Furthermore, an ICP-MS based method was also developed for quantitation of Tf in both serum and CSF samples using indium as a tracer. This novel approach took advantage of the metal-binding ability of Tf and the superb sensitivity of ICP-MS detection. Reliable quantitation of In_2Tf can be easily achieved without protein enrichment, thus providing an attractive alternative to LC-MS based approach. Although MS-based method is not a mainstay for quantitation of protein drugs currently, development of novel methods as discussed in this work will definitely broaden the application of mass spectrometry for tasks ranging from pharmacokinetic studies to quality control of biopharmaceutical products.

Table 1. Top five proteins identified by MASCOT MS/MS ion search engine

A. Without Ni-affinity purification (1 μg of Tf*, 1 mL of bovine serum)		
Protein List	Origin	Mascot Score
Albumin	Bos taurus	527
Transferrin	Bos taurus	472
alpha-2-HS-glycoprotein precursor	Bos taurus	251
Prothrombin	Bos taurus	101
apolipoprotein C-III precursor	Bos taurus	89
B. With Ni-affinity purification (1 μg of Tf*, 1 mL of bovine serum)		
Protein List	Origin	Mascot Score
Albumin	Bos taurus	585
hemopexin precursor	Bos taurus	496
Transferrin	Human	356
complement factor H precursor	Bos taurus	332
apolipoprotein C-III precursor	Bos taurus	321

Table 2. O^{18}-incorporation efficiency in cysteine-containing tryptic peptides of Tf*

Tryptic Peptides	Percentage of differentially O^{18}-incorporated peptides				
	$0xO^{18}$	$1xO^{18}$	$2xO^{18}$	$3xO^{18}$	$4xO^{18}$
KPVEEYANC*HLAR	<1%	11%	89%	N/A	N/A
WC*ALSHHER	<1%	12%	88%	N/A	N/A
DYELLC*LDGTR	<1%	13%	87%	N/A	N/A
C*LKDGAGDVAFVK	<1%	10%	90%	N/A	N/A
SVIPSDGPSVAC*VKK	<1%	11%	89%	N/A	N/A
DLLFRDDTVC*LAK	<1%	10%	90%	N/A	N/A
FDEFFSEGC*APGSK	<1%	10%	90%	N/A	N/A
WC*AVSEHEATK	<1%	13%	87%	N/A	N/A
C*STSSLLEAC*TFR	<1%	<1%	1.8%	17%	81%
EGTC*PEAPTDEC*KPVK	<1%	<1%	2.0%	18%	80%
IEC*VSAETTEDC*IAK	<1%	<1%	2.0%	17%	81%

Table 3. Validation of quantitation method of Tf* in rat CSF

low QC	theor conc (ng/50 µL)	number of replicates (n)	CV (%)	accuracy (%)
	60	5	10.2	90.3
Medium QC	600	5	4.75	104
High QC	1600	5	9.92	101
LLOQ	10	5	10.3	81.2
Matrix effect	500 ng/100 µL	3	4.93	104

Table 4. Validation of quantitation method of indium in rat CSF

QC samples	theor conc (ppb)	number of replicates (n)	CV, %	accuracy, %
Low	0.04	3	0.435	89.1
Medium	0.4	3	0.764	98.9
High	4	3	1.47	107
Low with 100 µL CSF	0.04	3	1.89	105
Medium with 100 µL CSF	0.4	3	0.975	106
High with 100 µL CSF	4	3	1.77	112

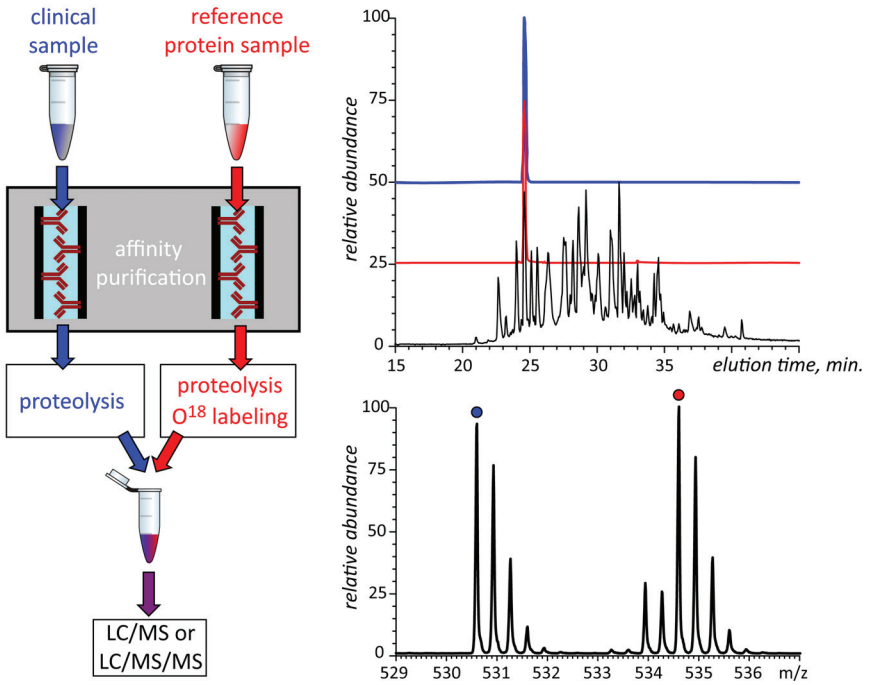


Figure 2-1. A general workflow of MS-based method for protein quantitation. Figure adapted from reference [77].

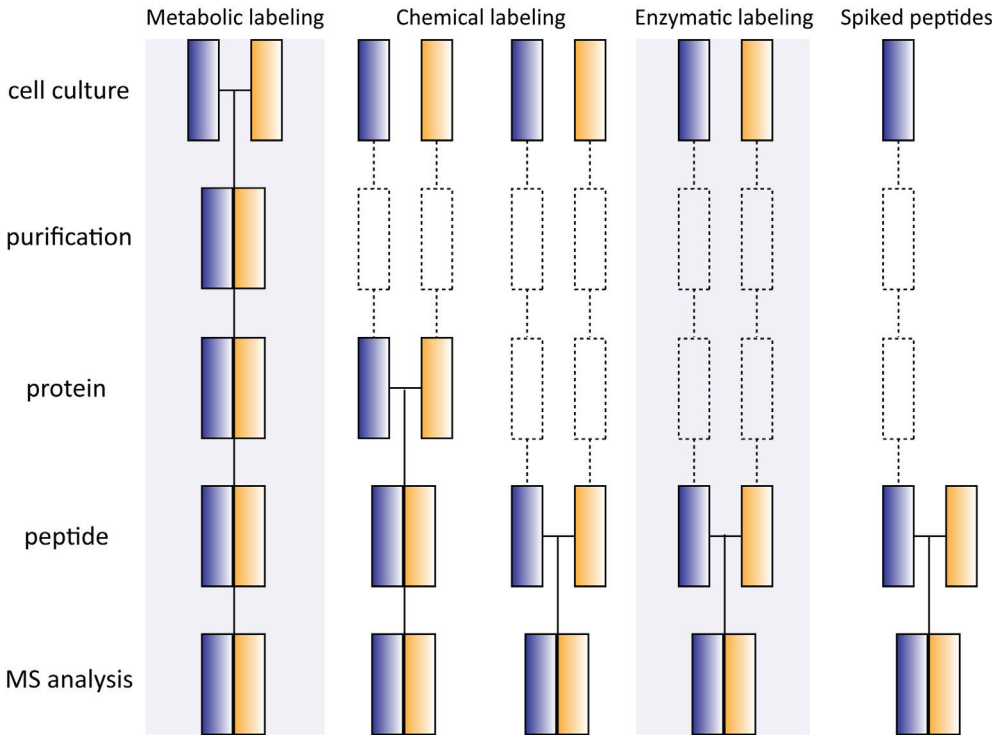


Figure 2-2. Strategies used for preparation of isotopically labeled internal standards. Blue: biological/clinical sample; yellow: internal standards. The horizontal line represents the stage where internal standards are spiked. Figure adapted from reference [31].

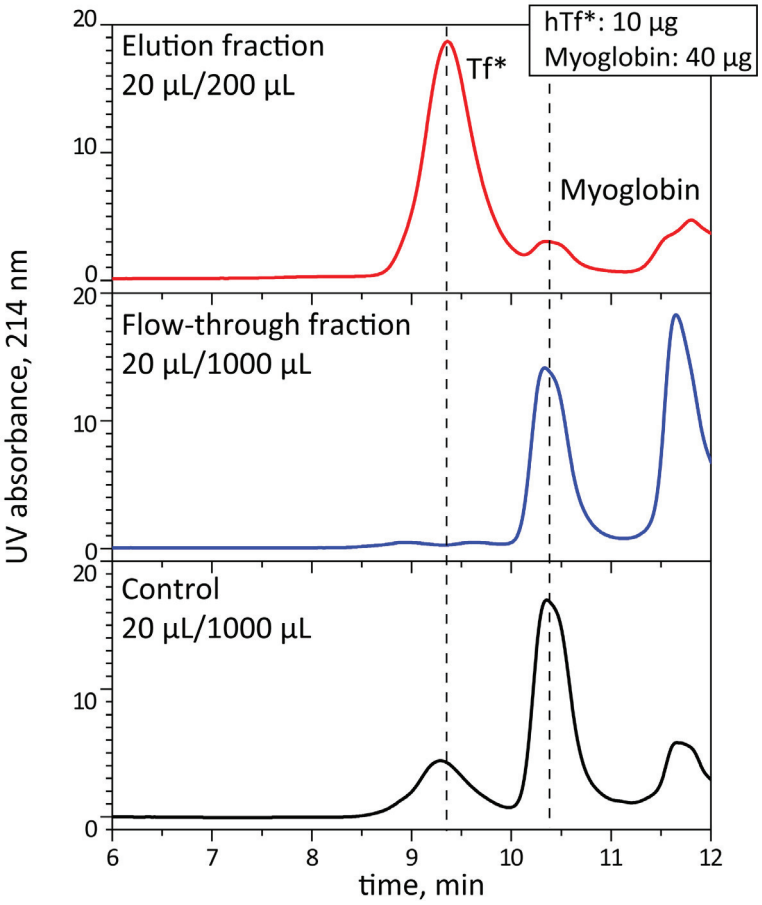


Figure 2-3. Investigate the recovery of Tf* after Ni-affinity purification by SEC. Black trace: control sample; blue trace: the first flow-through fraction; red trace: the elution fraction.

Figure 2-4. Investigation of Ni-affinity purification efficiency on bovine serum sample. Lanes: 1- NEB 2-212 kDa MW Ladder; 2- Tf*; 3- bovine serum; 4- Tf* + bovine serum; 5- flow through fraction; 6- purification fraction; 7- Tf*.

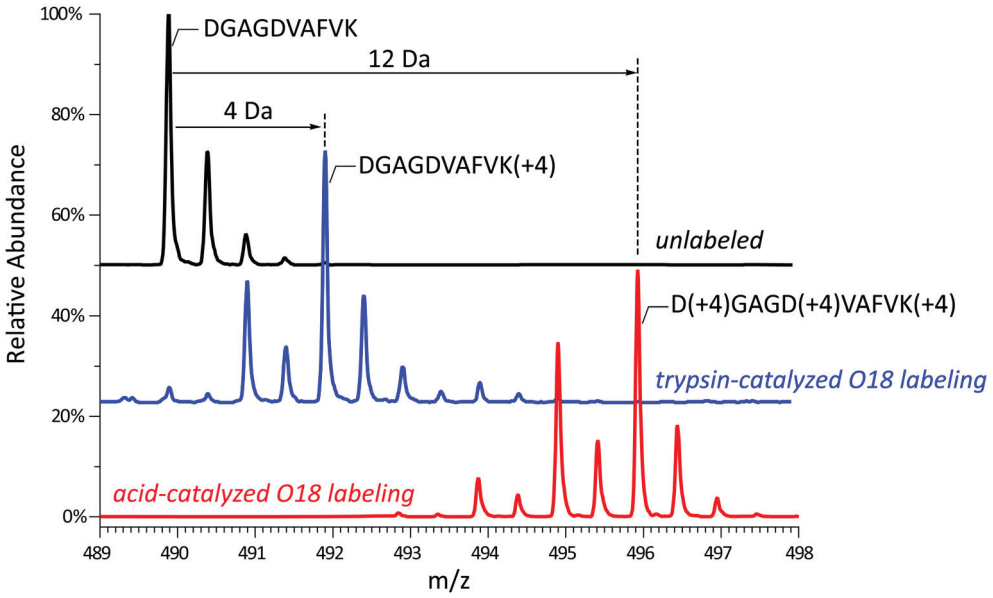


Figure 2-5. O18-labeled peptide DGAGDVAFVK prepared by trypsin-catalyzed O18 labeling (blue trace) and acid-catalyzed O18 labeling (red trace), in comparison with unlabeled peptide (black trace).

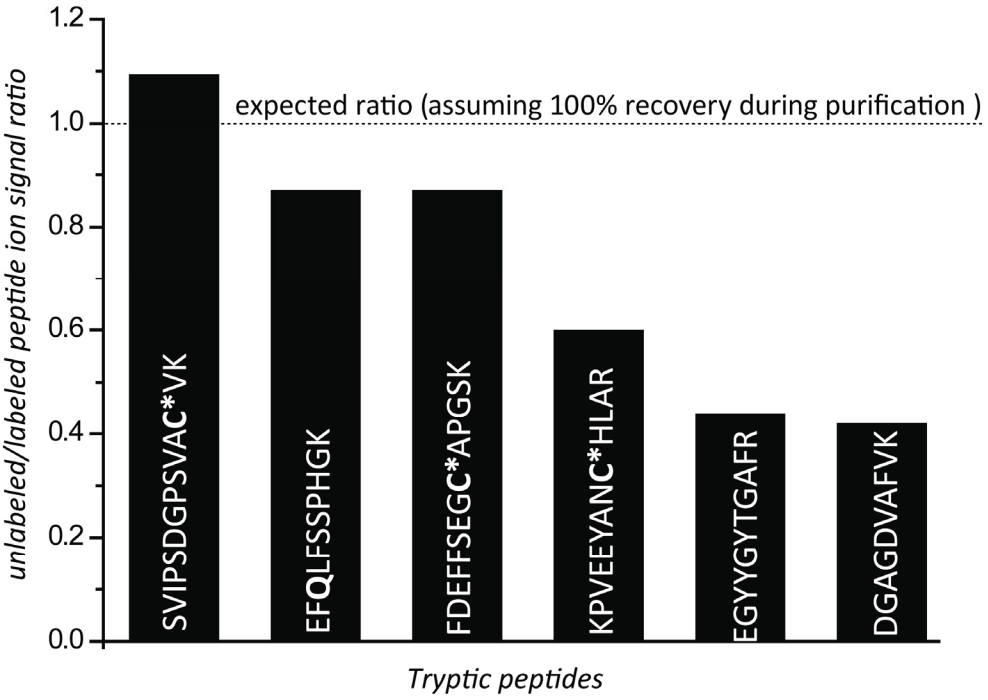


Figure 2-6. Quantitation of Tf* in bovine serum using different reference peptides. Identical amount of O18-labeled internal standard was added before LC-MS analysis.

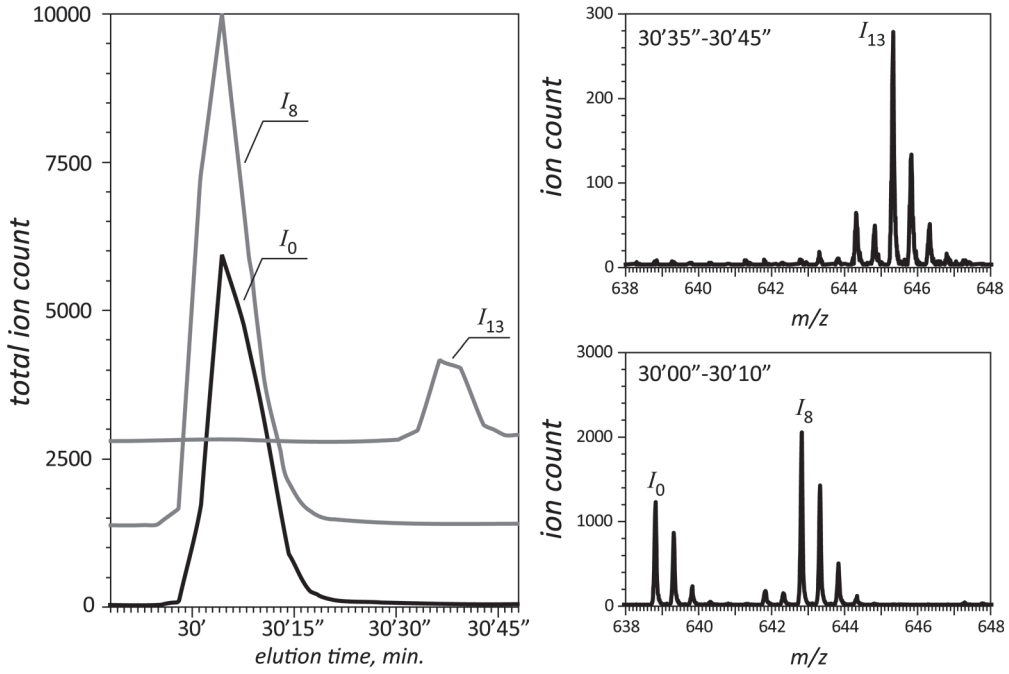


Figure 2-7. Acid-driven deamidation demonstrated by peptide EFQLFSSPHGK from tryptic digestion of Tf*. Left column represents extracted ion chromatograms (XICs) of ionic species I_0, I_8 and I_{13} respectively. Right column represents mass spectra at elution time 30'35"-30'45" and 30'00"-30'10" respectively.

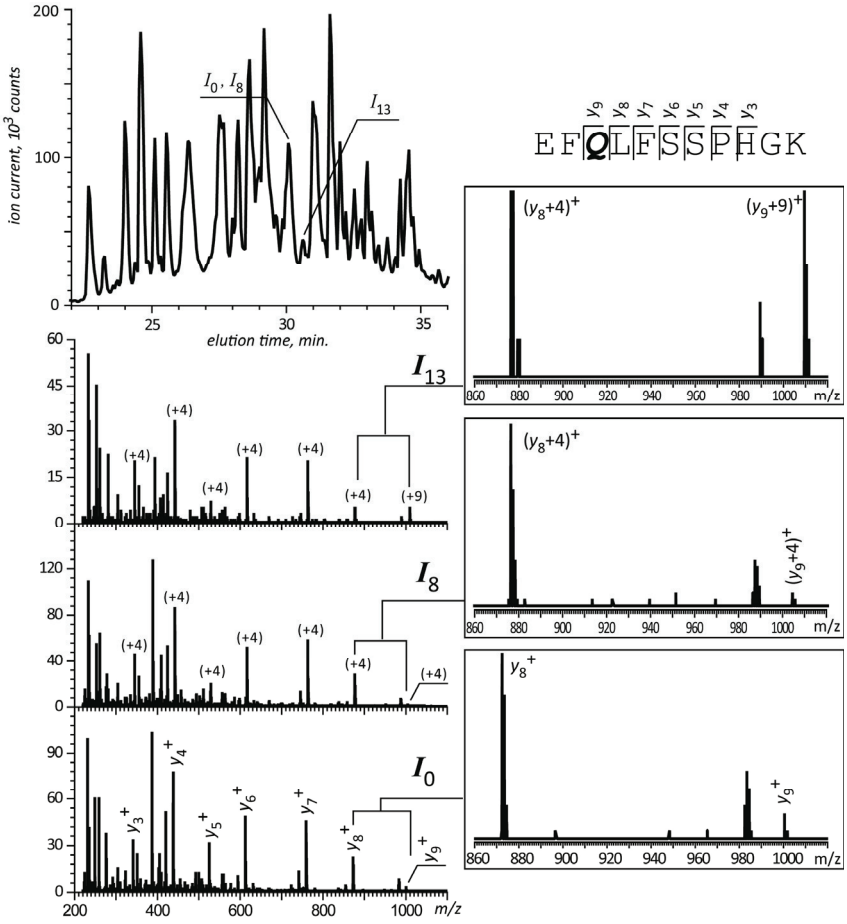


Figure 2-8. LC-MS/MS analysis of the structure of ionic species I_0, I_8, and I_{13} related to tryptic fragment EFQLFSSPHGK showing full views of fragment ions spectra (left column) and zoom views of m/z regions corresponding to y^{5+} and y^{6+} ions (right column). The retention times for the various parent ions are indicated in the total

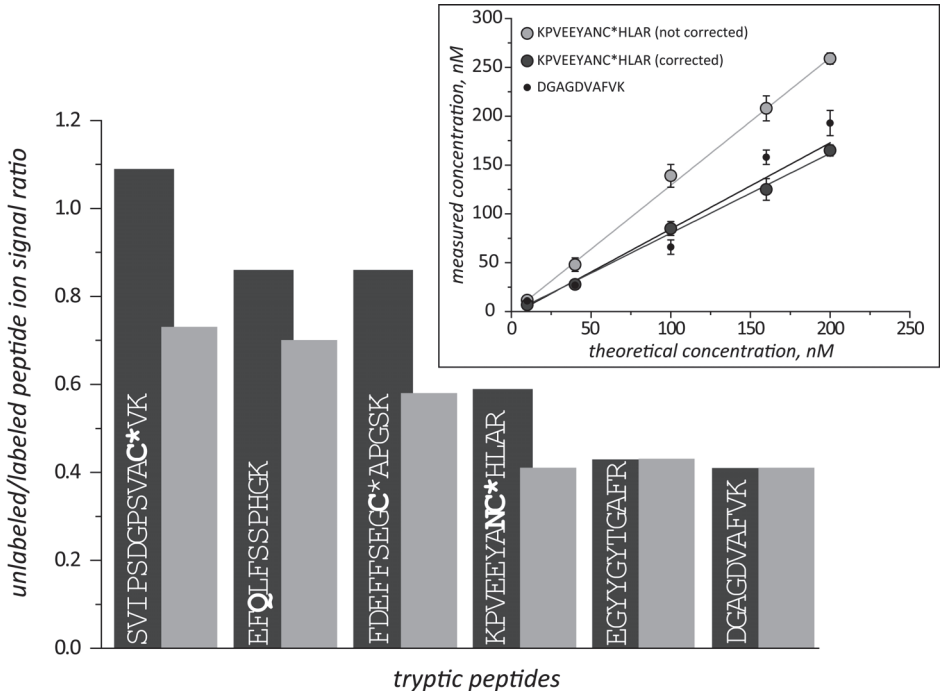


Figure 2-9. Quantitation of Tf* in bovine serum using different reference peptides. Identical amount of O18-labeled internal standard was added before LC-MS analysis. Results were shown both with (grey bars) and without (black bars) correcting for internal standard deamidation effect. The inset shows the calibration curves for two select peptides before and after correcting for internal

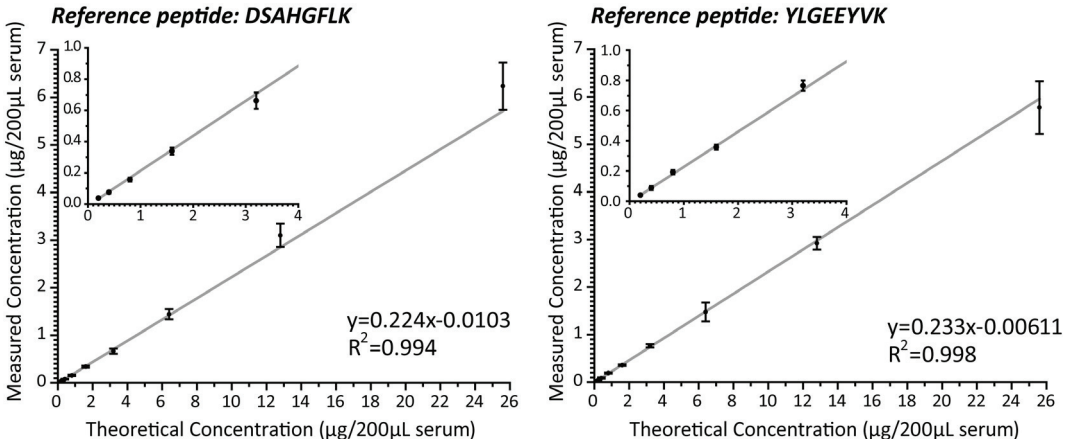


Figure 2-10. Calibration curves for quantitation of Tf* in bovine serum using reference peptides DSAHGFLK (left) and YLGEEYVK (right).

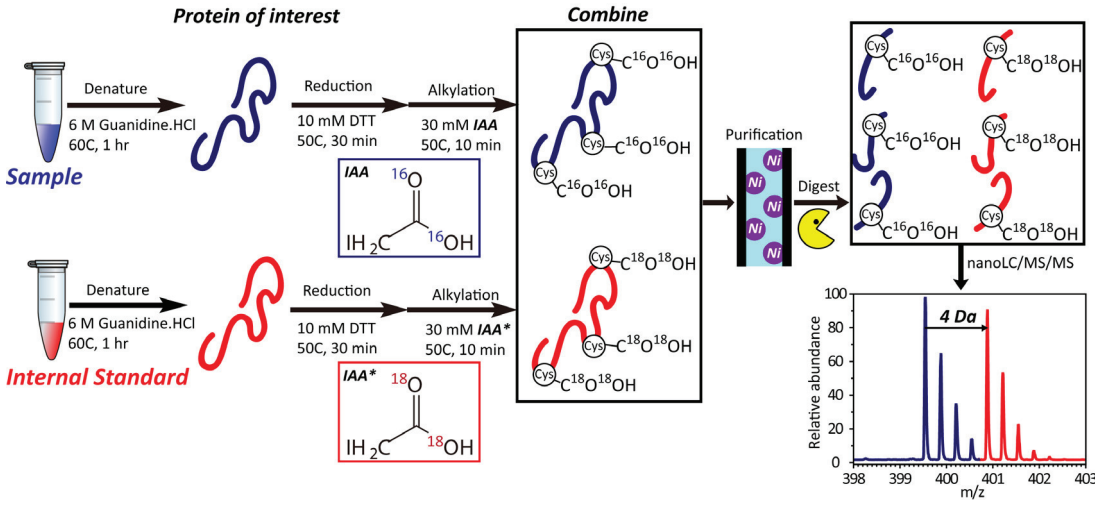


Figure 2-11. A schematic representation of applying Cys O18-labeling for accurate protein quantitation.

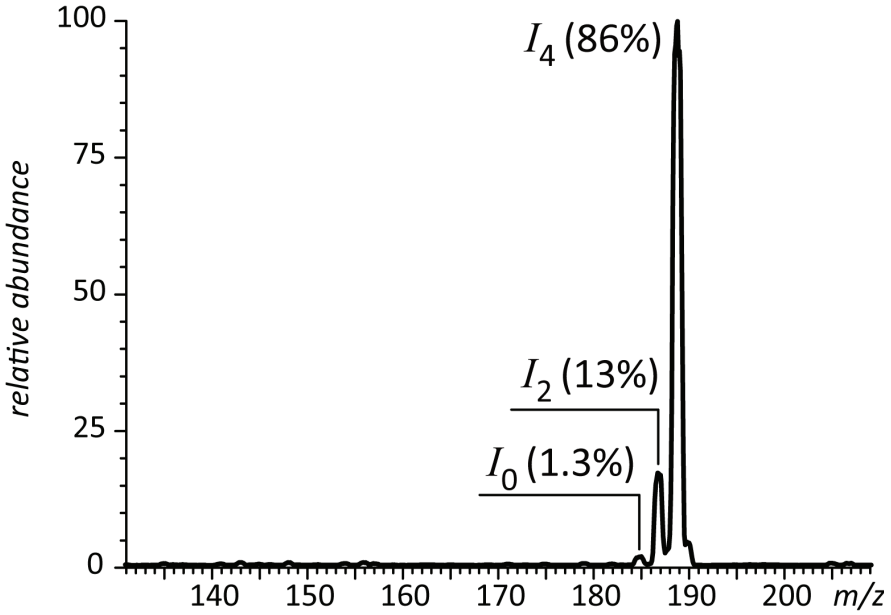


Figure 2-12. ESI MS analysis of O18-labeled IAA in the negative ion mode showing the isotopic distribution of IAA anion.

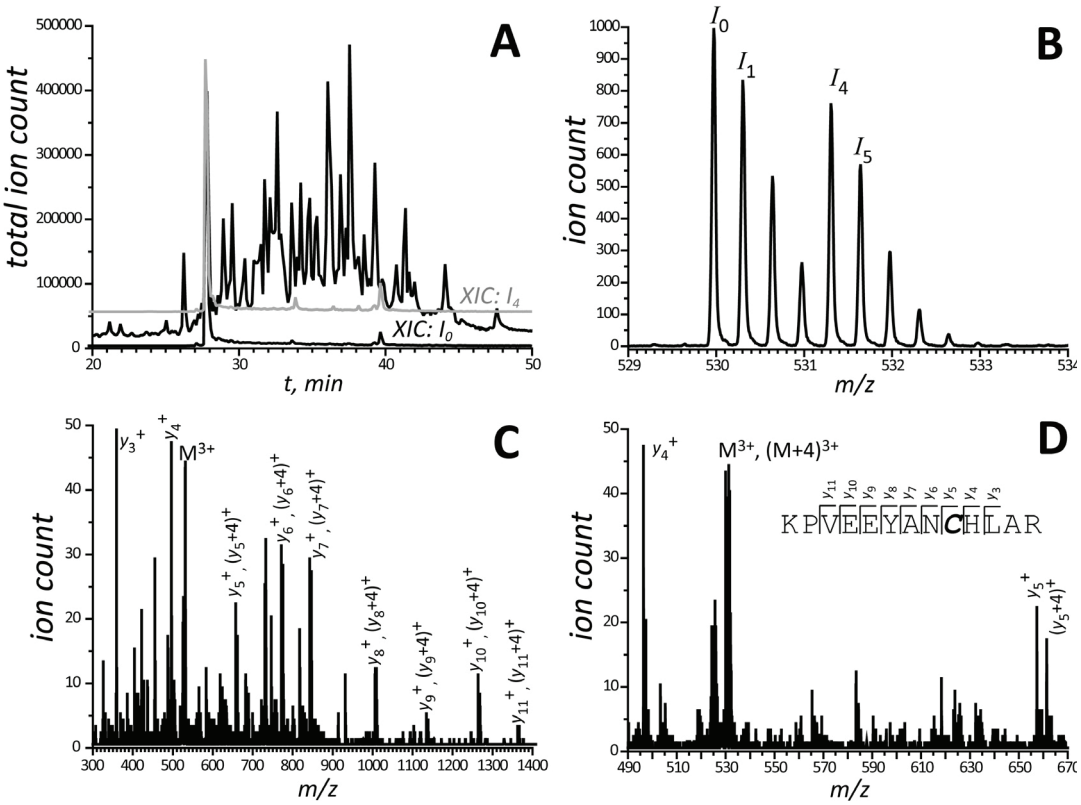


Figure 2-13. Nano-LC/MS/MS analysis of tryptic peptides derived from a mixture of hTf capped with unlabeled and O18-labeled IAA. (A) Total ion chromatogram of the digest and extracted ion chromatograms of peak I_0 and I_4 representing unlabeled (black trace) and O18-labeled (gray trace) tryptic fragment KPVEEYANC*HLAR. (B) Isotopic distribution of a triply charged peptide ion KPVEEYANC*HLAR. (C) MS/MS (CID) of peptide ion KPVEEYANC*HLAR shown in (B). (D) A zoom view of the MS/MS spectrum (c) showing y^{4+} and y^{5+} fragment ions

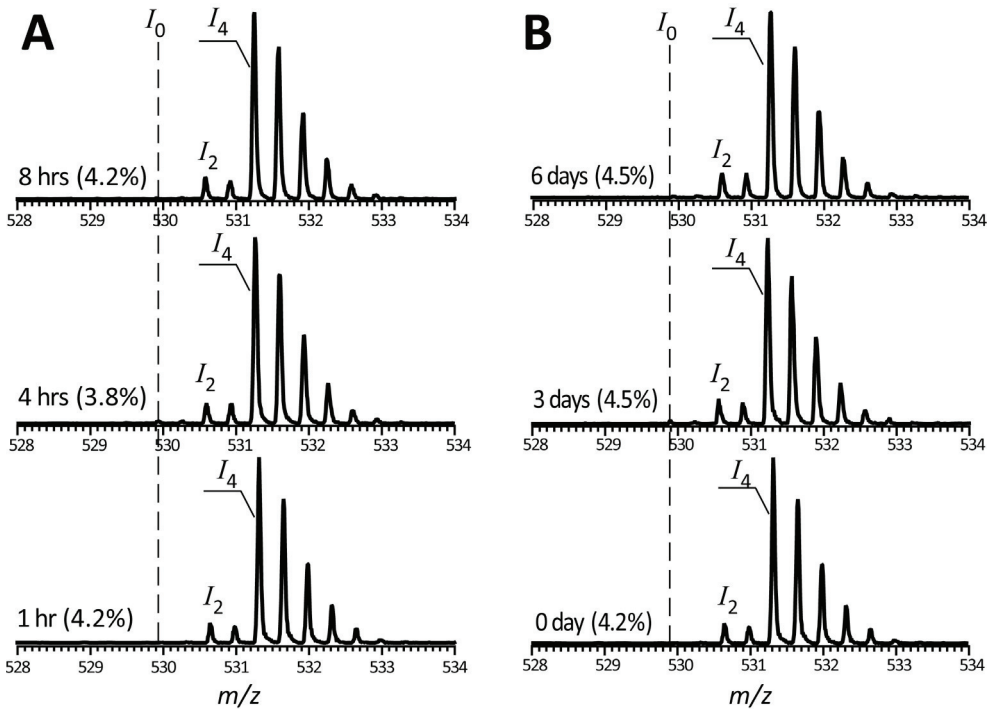


Figure 2-14. Stability of O18-labeled peptide KPVEEYANC*HLAR prepared by Cys O18-labeling under acidic condition (0.1% FA) (A) and near neutral pH (pH 8.0) (B). Number in the bracket represents the relative abundance of the isotopic peak I_2 as percentage of the total ionic signal.

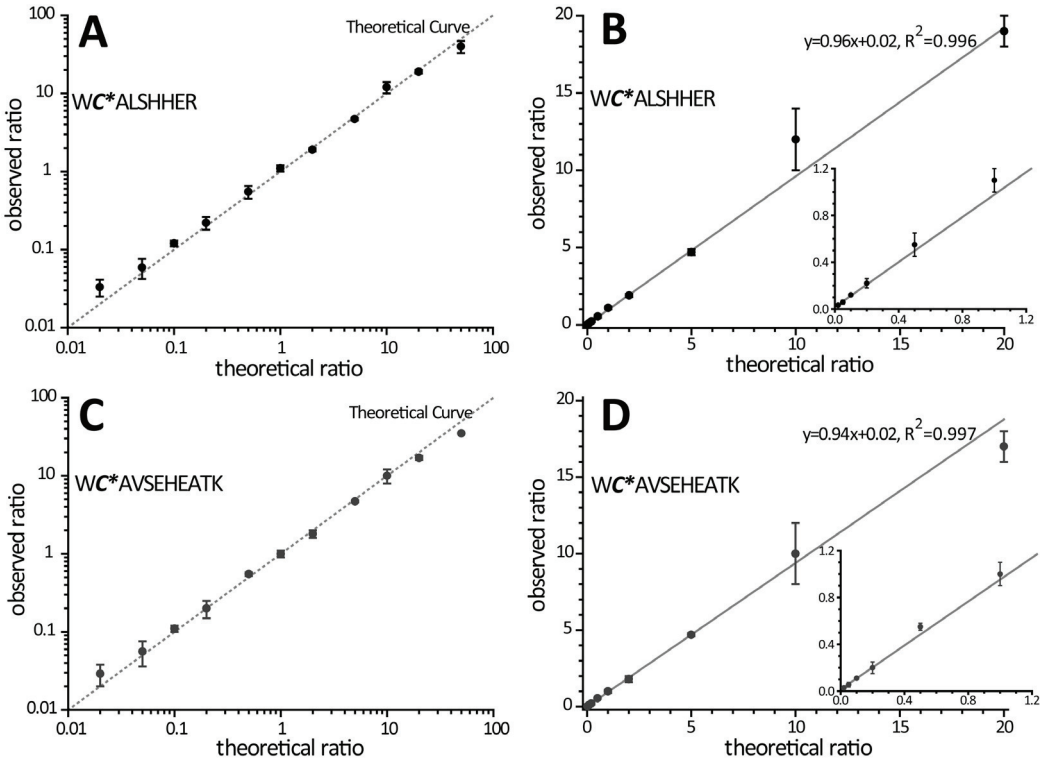


Figure 2-15. Quantitation curves for Tf* constructed using tryptic peptides WC*ALSHHER (A and B) and WC*AVSEHEATK (C and D) plotted on logarithmic and linear scales.

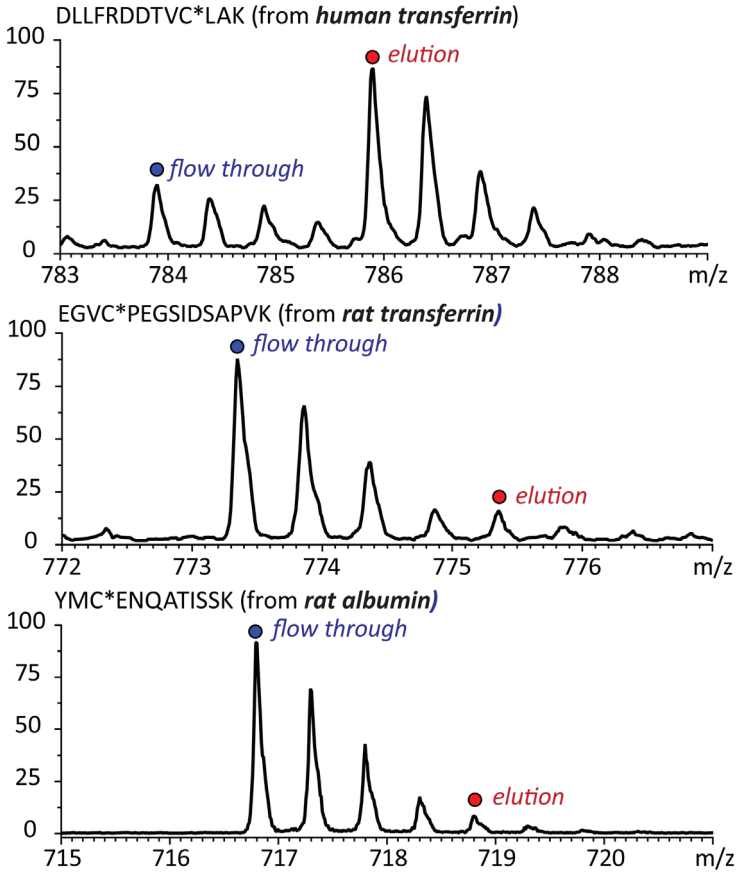
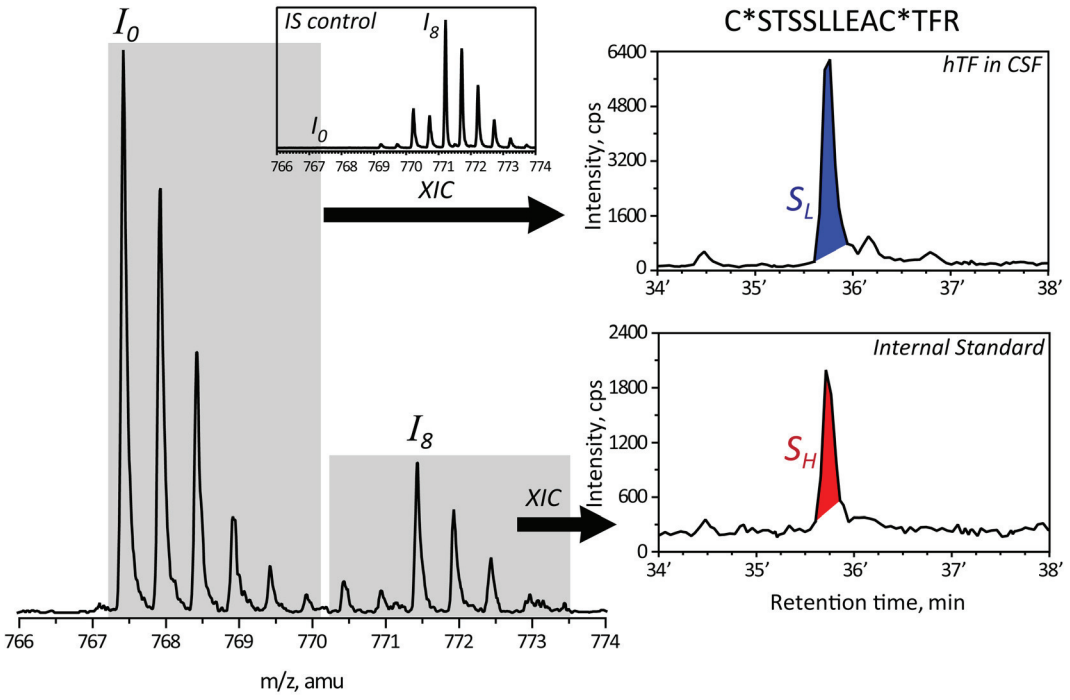
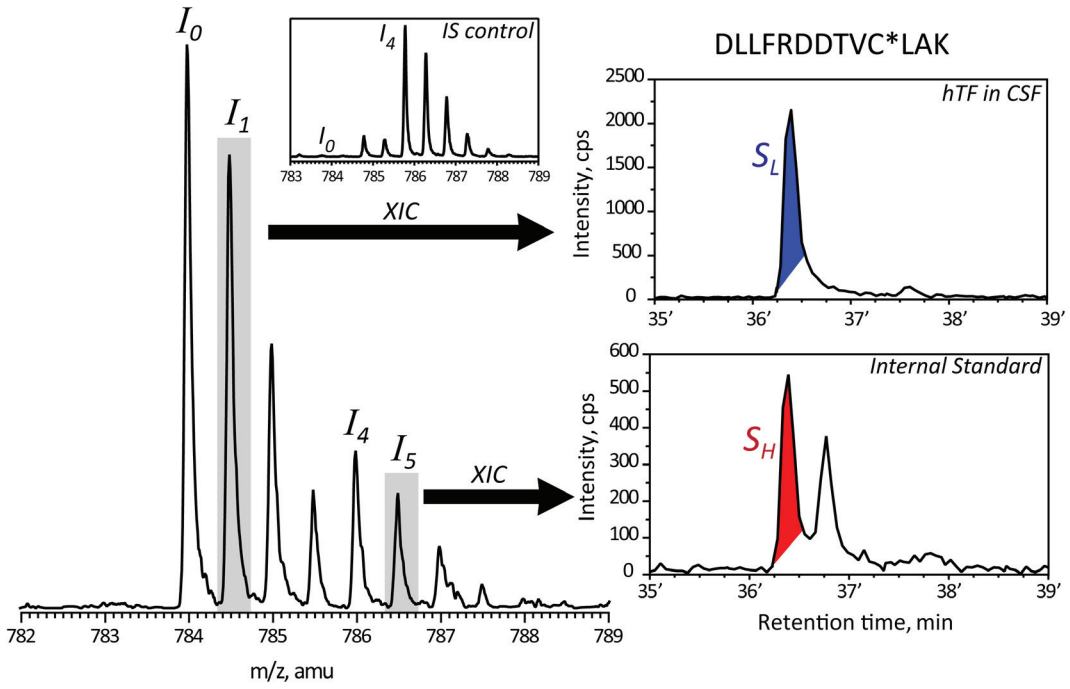


Figure 2-16. Evaluation of enrichment or removal of proteins from Tf*/rat CSF sample after Ni-affinity purification. Blue: flow-through fraction; red: elution



$$ratio = \frac{S_L}{S_H}$$, α (α=0.14 in this work) is the fraction of partially labeled peptides

Figure 2-17. Quantitation method using peptide containing two cysteine residues. The m/z range highlighted by grey area was extracted to generate XICs shown in the right column. The inset represents the control spectrum of this peptide alkylated by IAA-18.



$$ratio = \frac{S_L}{S_H \times (1+\alpha)}$$, α (α=0.14 in this work) is the fraction of partially labeled peptides

Figure 2-18. Quantitation method using peptide containing a single cysteine residue. The second isotopic peak of either unlabeled or fully labeled peptide highlighted by grey area was extracted to generate XICs shown in the right column. The inset represents the control spectrum of this peptide alkylated by IAA-18

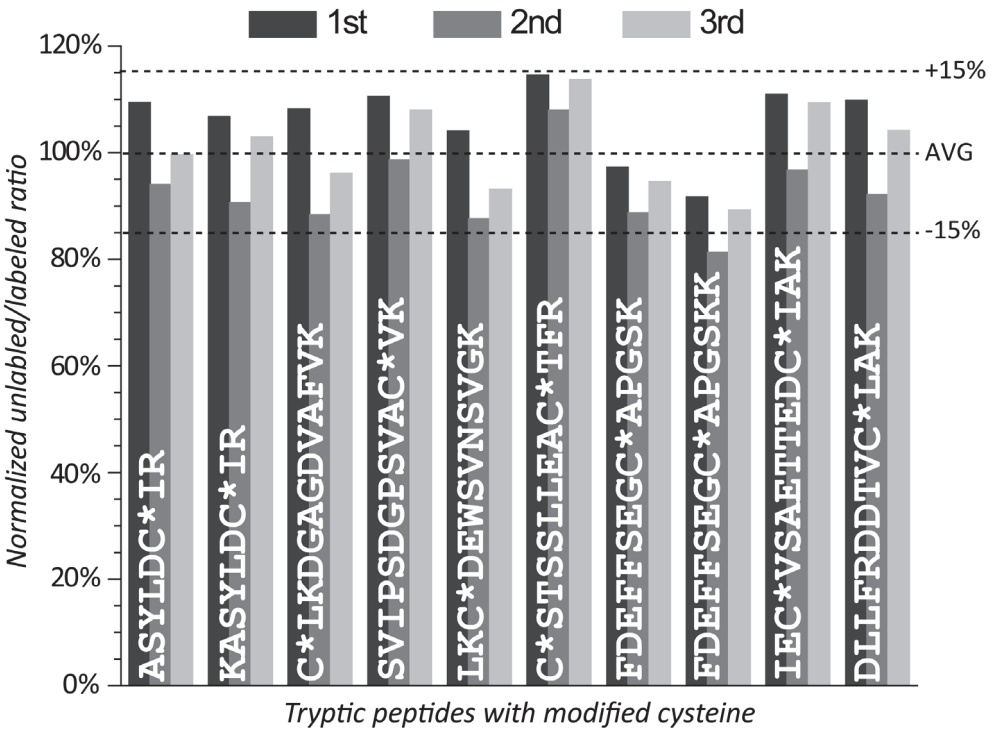


Figure 2-19. Comparison of quantitation results by different reference peptides. Quantitation results from triplicate experiments were shown in different bars.

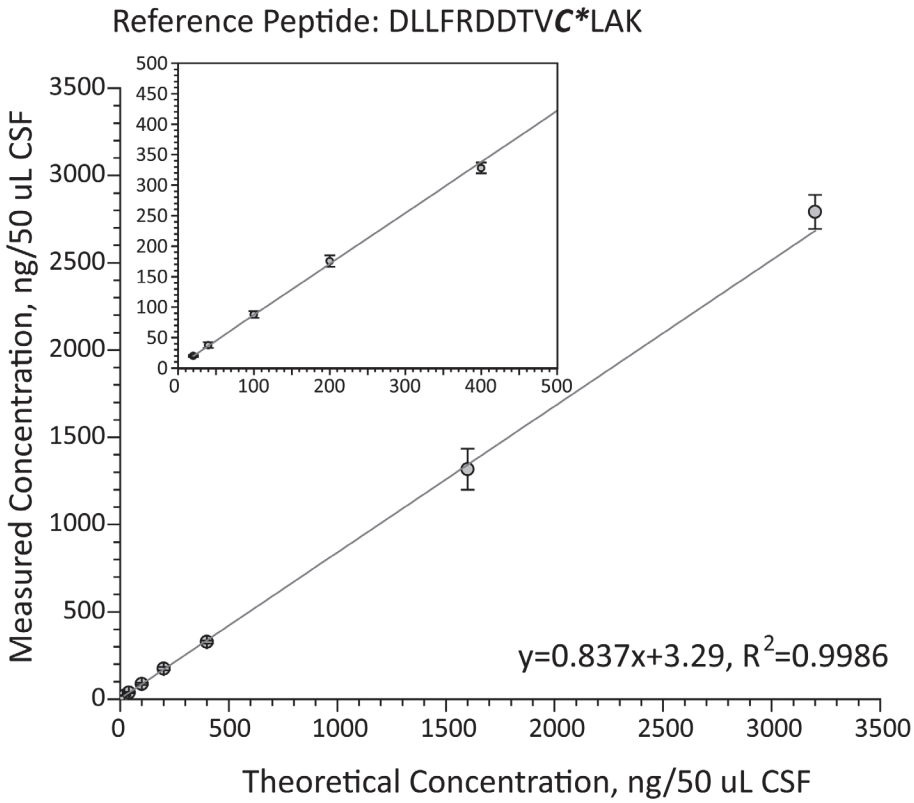


Figure 2-20. Calibration curve for quantitation of Tf* in 50 µL of rat CSF. A $1/X^2$ regression was applied. The inset shows an enlargement of the low end of the calibration curve.

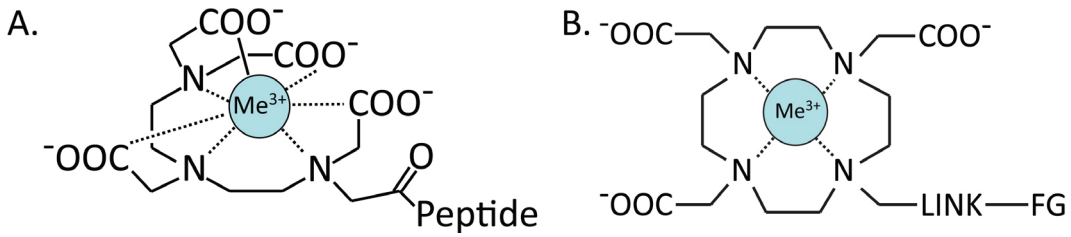


Figure 2-21. Structure of chelates DTPA (A) and DOTA (B). Me^{3+}: metal ion; FG: functional group

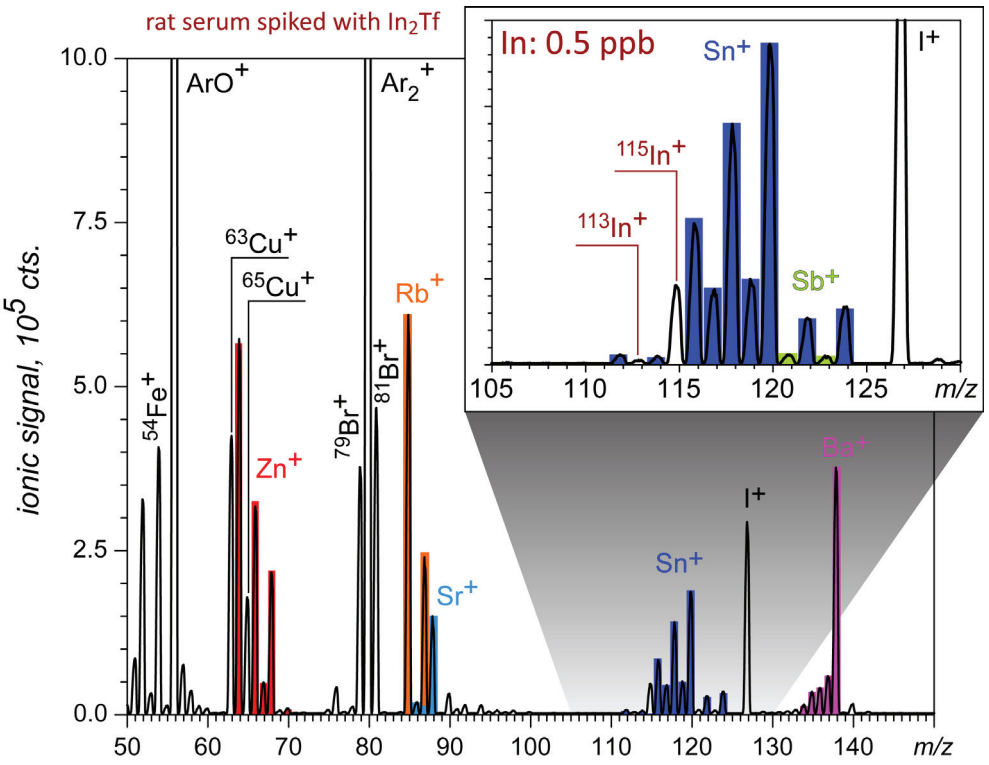


Figure 2-22. Full-scan spectrum of serum sample spiked with In_2Tf acquired by ICP-MS. Colored bars represent natural isotope distribution of several elements. The inset shows the signal from indium (both ^{113}In and ^{115}In) at a concentration of 0.5 ppb.

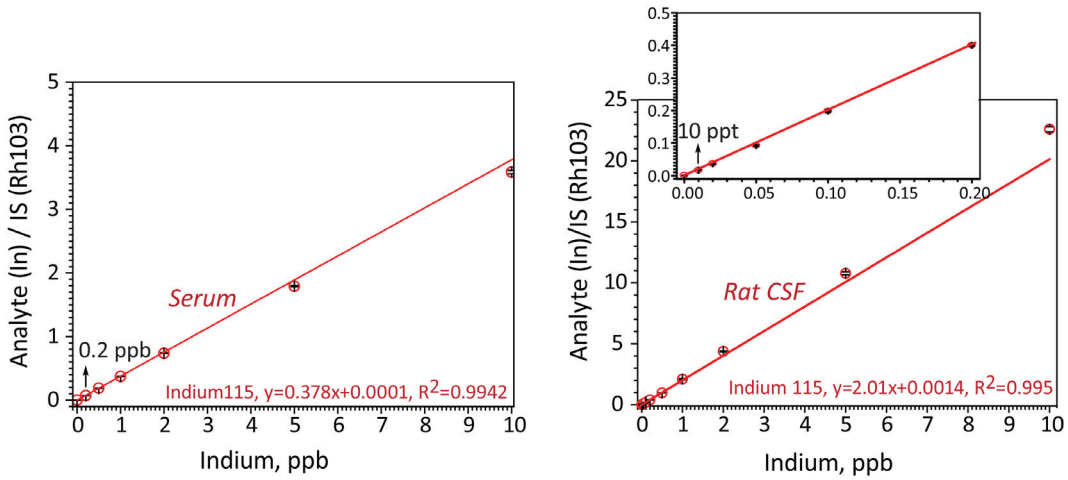


Figure 2-23. Calibration curves for quantitation of indium in serum (left) and rat CSF (right). The inset shows an enlargement of the low end of the calibration curve.

CHAPTER 3

O^{18}-LABELING BASED METHOD FOR IDENTIFICATION OF DEAMIDATION PRODUCTS

3.1 Introduction

As discussed in section 1, deamidation of asparagine has been observed to have a negative impact on protein quantitation. In fact, deamidation of asparagine and isomerization of aspartic acid in proteins are the most frequent non-enzymatic post-translational modifications (PTM) in *vivo* [66]. In a recent study, Liu *et al*. observed that endogenous IgG isolated from human serum possessed 23% deamidation on Asn384 [67]. Under physiological conditions, both deamidation and isomerization proceed through formation of a succinimide intermediate, and lead to a mixture of aspartyl (Asp) and isoaspartyl residues (isoAsp) [68]. Upon each deamidation event, an additional negative charge is introduced to the protein surface; furthermore, the formation of isoAsp residue elongates the protein backbone by inserting an extra methylene group, resulting in the β-peptide linkage. Thus, these modifications closely associate with protein conformation [69], function [70, 71], activity [72, 73], stability [74], aggregation [75] and even immune responses [76, 77]. *In vivo*, the level of isoAsp is controlled by protein L-isoaspartyl O-methyltransferase (PIMT), which catalyzes the conversion of isoAsp back to Asp [78]. The lack of such a mechanism *in vitro* leads to accumulation of isoAsp contents over time as Asn deamidation and Asp isomerization can also occur during protein production and long term storage. It is particularly

important for protein-based therapeutics, where the effect of isoAsp formation has been extensively examined [79]. Therefore, close monitoring of Asn and Asp degradation products in protein pharmaceuticals is highly desirable and necessary [80].

3.1.1 Traditional methods for identification of deamidation products

Traditional ways of detection and localizing isoAsp residues in stressed protein involve chemical hydrolysis and proteolysis. Edman sequencing is useful for identifying the formation of the isoAsp residue, because these reactions stop at the iso-peptide bond [81]. PIMT is also frequently applied, since it leads to the methylation of isoAsp residues as an intermediate step. Owing to the longer retention time of methylated isoaspartyl-containing peptide compared to its aspartyl counterpart on RP-HPLC analysis, valid assignment can be made [82]. Furthermore, the byproduct of the methylation reaction, S-Adenosyl-L-homocysteine (SAH), has been used to achieve global analysis of isoAsp residue content in protein samples, which has already been commercialized under the name IsoQuant [83, 84]. The utility of endoproteinase Asp-N for the differentiation of aspartyl- and isoaspartyl-containing peptides has also been explored, by taking advantage of the selective cleavage of aspartyl-containing peptides but not their isoaspartyl counterparts [85, 86].

3.1.2 MS and LC-MS based methods for characterization of deamidation products

Detection of deamidated peptides by MS is relatively straightforward, since a mass shift of +0.984 Da occurs upon each deamidation event, which can be readily detected

by MS with high resolution. However, the differentiation between aspartyl- and isoaspartyl-containing peptides from Asn deamidation or Asp isomerization presents a significant challenge for MS due to their identical mass. Even when separated by HPLC, unambiguous identification of the two isomers can be difficult using LC-MS analysis alone. In fact, most of the assignments have been made based on either the ca. 3:1 relative peak intensity ratio or the elution order of the aspartyl- and isoaspartyl-containing peptides (isoaspartyl-containing peptides are believed to be more hydrophilic than their aspartyl counterparts due to the greater acidity at the side-chain, therefore have shorter retention time during RP-HPLC analysis). However, as has been reported by other researchers [86-88], the elution order is strongly dependant on the chromatographic conditions as well as the properties of the peptides. In fact, the inverted elution order (isoaspartyl-containing peptides have longer retention time than aspartyl-containing peptides) has been consistently observed in peptides with Asp/isoAsp residues located at the non-acetylated N-terminus [87, 89]. We have also observed examples with inverted elution order in this work (vide infra). Clearly, accurate assignment of aspartyl- and isoaspartyl-containing peptides cannot be achieved by elution order alone. Likewise, assignment of aspartyl- and isoaspartyl-containing peptides based on relative peak intensity ratio (ca. 1:3) is not always reliable. Since these peptides usually elute at different time, their relative ionization efficiencies can be modulated differently by mobile phase and/or co-eluting species, and may not reflect their fractional concentrations in solution. Furthermore, the 1:3 ratio had been

established based on the analysis of products of deamidation of short, unstructured peptides, while in proteins it may vary due to influence of the higher order structure [90, 91].

3.1.3 Tandem MS-based methods for differentiation of deamidation products

Other approaches to aspartyl- and isoaspartyl-containing peptides differentiation have been mainly focused on applying different tandem mass spectrometry techniques. Several successful examples have been reported using different collision-activated dissociation (CAD) techniques, in which the differentiation is mainly based on the specific reporter ions [92] or the relative fragment ion intensity ratios [93]. However, these approaches are highly dependent on the peptide sequence and very sensitive to experimental conditions. A much more successful strategy is the application of electron-based fragmentation techniques (electron capture dissociation, ECD, and electron transfer dissociation, ETD), which generate a pair of reporter ions ($c^{\cdot}$ +57 and z-57) that are unique to isoaspartyl-containing peptides [94], and another pair of reporter ions (z-44 and $(M+nH)(n-1)^{+\cdot}$ -60) that are unique to aspartyl-containing peptides [95, 96]. This approach has already been applied in a high-throughput manner to achieve proteome-scale identification and quantitation of isoaspartyl residues in biological samples [97, 98]. However, this technique is limited to ECD-capable instruments, and sometimes can be hampered due to the low abundance of the reporter ions [86].

3.1.4 An O^{18}-labeling approach to differentiation of deamidation products

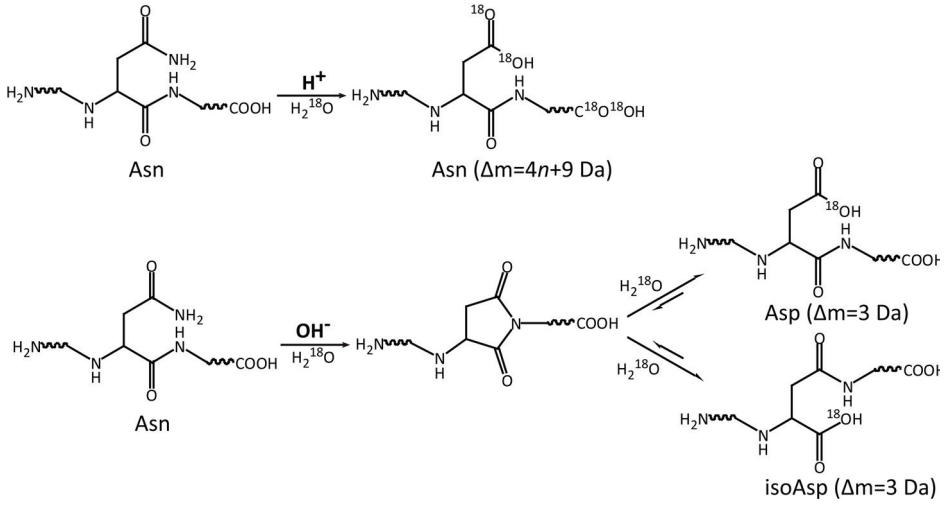
Recently, a number of O^{18}-labeling based strategies were developed to facilitate the Asn deamidation or Asp isomerization studies [99, 100]. Most of them have taken advantage of the increased mass increment after deamidation due to O^{18} incorporation (from 1 Da to 3 Da). Particularly, by using H_2O^{18} for sample storage and protein digestion, *in vitro* and *in vivo* occurring deamidation processes can be readily differentiated [101, 102]. However, none of these methods have explored the ability of using O^{18}-labeling for aspartyl- and isoaspartyl-containing peptides differentiation and assignment. In the current work, we demonstrate the possibility of using O^{18}-labeling and LC-MS method for unambiguous differentiation of aspartyl- and isoaspartyl-containing peptides. Taking advantages of different deamidation mechanisms in acidic and basic environments, isomer-specific mass tags have been introduced to O^{18}-labeled aspartyl- and isoaspartyl-containing peptides, which then can be easily distinguished by MS. Feasibility of this new method has first been demonstrated using a synthesized peptide EWSVNSVGK. Subsequently, successful assignment was achieved for at least six Asn-containing peptides and two Asp-Gly motif-containing peptides, from the tryptic digestion of an 80 kDa non-glycosylated form of human serum transferrin (hTf), which is a part of a number of biopharmaceutical products that are currently under development [80, 103].

3.2. Results and discussion

3.2.1 O^{18}-labeling assisted approach to differentiation of deamidation products

The mechanisms of Asn deamidation under both basic conditions and acidic conditions have been extensively studied [68]. Unlike deamidation under basic conditions, which generates both aspartyl- and isoaspartyl-containing peptides, acid-catalyzed deamidation only leads to formation of aspartyl-containing peptides. By preparing forced deamidation standards in H_2O^{18} under different conditions, different levels of O^{18} incorporation in aspartyl- and isoaspartyl-containing peptides can be achieved, which can be exploited for unambiguous isomer differentiation (Scheme 4).

Scheme 4



As previously reported [56], acid-catalyzed labeling and deamidation result in O^{18} incorporation in all carboxylic groups, leading to a maximum mass increment of $4n + 9$ Da (4 Da from each acidic residues, 4 Da from C-terminus, and 5 Da from deamidation-formed Asp residue), where n is the number of acidic residues and

carboxymethylated Cys residues in this peptide. On the other hand, base-catalyzed deamidation only leads to a mass increment of 3 Da (1 Da from deamidation and 2 Da from incorporation of one O^{18} atom), because only one O^{18} atom is incorporated during the hydrolysis of succinimide intermediate. Indeed, due to the much lower rate of Asp/isoAsp isomerization under basic conditions, the formation and subsequent hydrolysis of succinimide intermediate for the second time is limited within the experimental time-scale.

3.2.2 Preparation of O^{18}-labeled deamidation standards

Under basic conditions, Asn deamidation in proteins occurs at a much lower rate compared to deamidation of short unstructured peptides due to the protection afforded by the higher order structures. Therefore, both acid- and base-catalyzed deamidation standards were prepared using peptides as starting material. Some variation in the deamidation rate under basic conditions was still observed, since this process is influenced by the sequence of the peptide, especially the neighboring residue on the C-terminal side of Asn. In particular, deamidation occurs much faster in Asn followed by a polar residue with a relatively small side chain (Gly, Ser, Thr and Asp), compared to that followed by a hydrophobic residue with a bulky side chain (Val, Leu and Ile) [104]. In order to accelerate the deamidation process, elevated pH (1% NH_4OH, pH ~ 10) and temperature (50 °C) were applied. In contrast, the deamidation rate of Asn under acidic conditions is barely affected by the peptide sequence. By incubating the model peptide

EWSVNSVGK in H_2O^{18} at 50 °C and in the presence of 1% TFA, the deamidation process was monitored by LC/MS analysis overtime. After 24 hours' incubation, nearly 15% of peptides were in aspartyl form, and the isotopic peak presenting maximum O^{18} incorporation could be readily detected (Figure 3-1). Thus, both acid- and base-catalyzed deamidation standards were prepared by 24 hours' incubation using the conditions described above.

3.2.3 A model peptide study

Forced deamidation of model peptide EWSVNSVGK was carried out in H_2O^{18} under both acidic and basic conditions using the procedure described above. Subsequently, the O^{18}-labeled deamidation standards were analyzed by LC-MS. As previously reported [56, 105], oxygen atom exchange occurs between carboxylic groups and solvent at low pH, leading to a maximum mass increment of 8 Da (I_8) for peptide EWSVNSVGK. In addition, hydrolysis of the amide group in Asn residue under acidic conditions leads to an additional mass increase of up to 5 Da (the I_{13} species). Indeed, only one deamidation product with a maximum mass increase of 13 Da (designated as peak 3 in Figure 3-2A) is observed under acidic conditions. In contrast, the base-catalyzed deamidation leads to the formation of two indistinguishable deamidation products (designated as peaks 1 and 3 in Figure 3-2B), with a maximum mass increase of 3 Da (the I_3 species) for each. Analysis of the mixture of acid- and base-catalyzed deamidation standards reveals a co-elution peak of species I_3 and I_{13} (peak 3) which can be unambiguously assigned to

aspartyl-containing peptide. The other chromatographic peak for species I_3 (peak 1) can be assigned to isoaspartyl-containing peptide (Figure 3-2C). The corresponding mass spectra also allow the unambiguous differentiation between the two isomers to be made.

3.2.4 Method validation by ECD-based approach

To validate this approach, a mixture of acid- and base-catalyzed deamidation standards prepared from the model peptide EWSVNSVGK was further analyzed by LC-MS/MS using ECD to produce fragment ions. ECD MS/MS analysis allows a distinction to be made between aspartyl- and isoaspartyl-containing peptides, since it can generate a pair of reporter ions ($c_r^{\cdot}$+58 and z_{k-r}-57) that are unique to isoaspartyl-containing peptides, and another pair of ions (z_{k-r}-44 and $(M+nH-60)^{(n-1)+\cdot}$) that are unique to aspartyl-containing peptides (r is the position of Asp/isoAsp residue from the N terminus and k is the total number of residues in the peptide). Both ion I_3 and ion I_{13} (elution at t_1 and t_3) were selected for ECD MS/MS analysis (Figure 3-3). The reporter ion (z_5-57) was only observed from the fragmentation of ion I_3 at elution time t_1, which confirmed the identity of this species as base-catalyzed isoaspartyl-containing peptide. On the other hand, both (z_5-44) and $(M+2H-60)^{+\cdot}$ fragment ions were observed in the ECD spectrum of species I_3 acquired at elution time t_3, but not from the fragmentation of species I_3 at elution time t_1, which confirmed the identity of this species as base-catalyzed aspartyl-containing peptide. In addition, the corresponding ions (z_5+4)-44 and

(M+2H+8)-60 with additional incorporation of O^{18} atoms were observed in the fragmentation of ion I_{13}, which confirmed the identity of this species as acid-catalyzed aspartyl-containing peptide. Despite the O^{18} incorporation in the carboxylic group of Asp residue, the reporter ions did not carry any O^{18} atom due to the consequent loss of O^{18} atoms in the form of either CO_2 (44 Da) or $C_2H_4O_2$ (60 Da) during the fragmentation. These results are in excellent agreement with the proposed mechanism [94] for the neutral loss involving Asp residues during ECD fragmentation. The ECD MS/MS analysis clearly validates the O^{18}-labeling approach to identification of aspartyl- and isoaspartyl-containing peptides in deamidated peptides.

3.2.5 Characterization of deamidation products from human Tf

Following validation of the new methodology with ECD MS/MS, it was applied to detect and identify deamidation products of an 80 kDa plasma protein hTf, whose ability to traverse physiological barriers and be internalized by malignant cells is attracting significant attention in the biopharmaceutical sector [80]. Deamidation of protein pharmaceuticals is always a big concern during the protein production and storage. A general procedure to monitor the deamidation in proteins is by detecting the deamidated peptides in the protein digest (mostly by trypsin) mixtures. However, as has been reported before [88], the traditional overnight tryptic digestion can introduce false positives (artificial deamidation, especially in Asn-Gly motif). Therefore, in this study, a shortened digestion time (4 hours) was used to minimize the digestion-induced

deamidation. Accordingly, an extra step of removing the alkylating reagent as well as an increased enzyme to substrate ratio was applied in order to improve the digestion efficiency. The tryptic digests were further subjected to acid- and base-catalyzed deamidation in H_2O^{18}, and followed by LC/MS analysis.

Successful assignment of aspartyl- and isoaspartyl-containing peptides was achieved for six Asn-containing peptides and two Asp-Gly motif containing peptides from hTf (Figure 3-4). Based on the number of acidic residues in the peptide (*n*), a specific ion (I_{4n+9}) was extracted together with ion I_3. As discussed earlier, the co-eluting species I_{4n+9} and I_3 indicated the presence of aspartyl-containing peptide, while the other elution peak of ion I_3 indicated the presence of isoaspartyl-containing peptide. In the case of peptide DSGFQMNQLR, due to the presence of two Gln residues which are also targets of acid-catalyzed deamidation, three peaks of ion I_{13} eluting at different times were observed. Nevertheless, even in this case the assignment of aspartyl-containing peptide can still be made based on the co-elution of one of such peaks with one of the elution peaks of ion I_3. By subtracting, the other two peaks of ion I_{13} can be assigned to either one of the Gln-deamidated peptides. Furthermore, by extracting ion I_2, the isoaspartyl-containing peptide produced by Asp isomerization can be easily detected. Since Asp isomerization occurs much slower than Asn deamidation under basic conditions, it was only observed in peptides with an Asp-Gly motif which is the most susceptible isomerization site.

3.2.6 Elution order of aspartyl- and isoaspartyl-containing peptides in RP-HPLC

The elution order of aspartyl- and isoaspartyl-containing peptides from tryptic digestion of hTf was mostly in agreement with the general observation that isoaspartyl-containing peptides elute earlier than aspartyl-containing peptides due to the greater acidity at the side-chain. However, the inverted elution order did occur in the case of peptides NLREGTCPEAPTDECKPVK and DGAGDVAFVK (Figure 3-4), where both the Asn residue and Asp-Gly motif were localized at the N-terminus. It is likely that the more favorable interaction between the N-terminal amine group and the carboxyl group of isoAsp residue could increase the overall hydrophobicity of the peptide, leading to longer elution time (Figure 3-5). Indeed, another peptide CLKDGAGDVAFVK, as an extended version of peptide DGAGDVAFVK with three more residues at the N-terminal side of Asp, exhibited 'normal' elution order *vis-a-vis* aspartyl- and isoaspartyl-containing peptides (Figure 3-5). This could be explained by the unfavorable ionic interaction at N-terminus. Furthermore, peptide SVIPSDGPSVACVKK (Figure 3-5) also exhibited inverted elution order after Asp isomerization, probably caused by the hydrogen bonding between the side-chains of Ser and isoAsp residues. These results indicate the unreliability of assigning the aspartyl- and isoaspartyl-containing peptides based on elution order alone.

3.3 Conclusion

A new method based on O^{18}-labeling and LC-MS detection was developed for unambiguous assignment of aspartyl- and isoaspartyl-containing peptides produced by Asn deamidation and Asp isomerization. By preparing the acid- and base–catalyzed deamidation standards in H_2O^{18}, specific mass tags were introduced to aspartyl- and isoaspartyl-containing peptides respectively, which could be easily distinguished by MS. Compared to the traditional method where the assignment of the isomers is based on the assumption of elution order, this new method is both more accurate and reliable. Furthermore, since the method can be applied to the entire digest of a protein, multiple assignments can be made simultaneously. The procedure is considerably more cost-effective and faster than the use of synthetic aspartyl- and isoaspartyl-containing peptide standards. The new method also offers an attractive alternative to the direct ECD MS/MS analysis of deamidation products, which relies on detection of reporter ions and, therefore, may not produce a conclusive answer if the fragment ion intensity is insufficient. Furthermore, isomer identification can be carried out with the new methodology using a generic and inexpensive LC-MS system without tandem capabilities, such as single quadrupole MS.

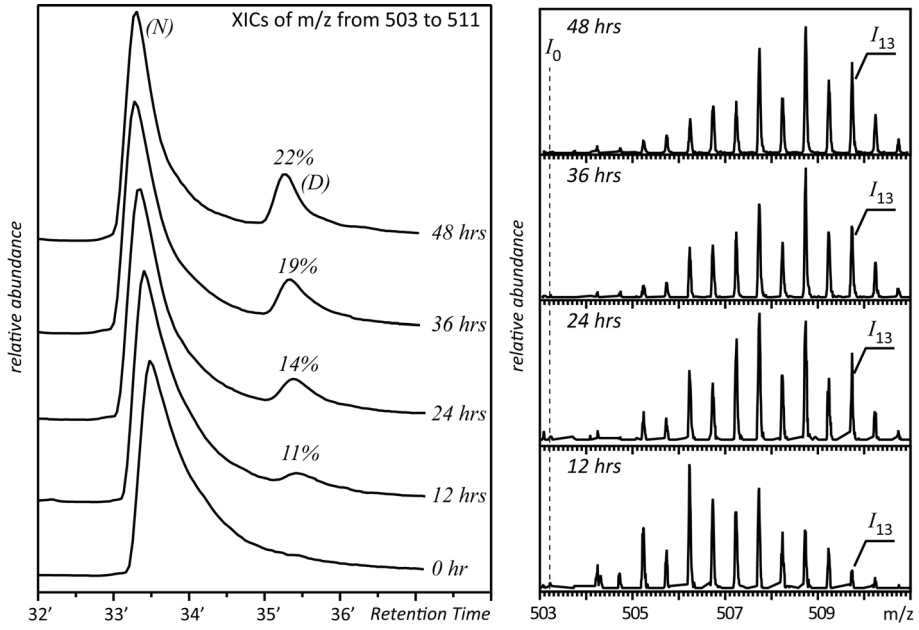


Figure 3-1. Process of acid-catalyzed deamidation (at 50 °C with 1% TFA) of model peptide EWSVNSVGK in H_2O^{18}. XICs of m/z range from 503 to 511 were generated to estimate the relative abundance of deamidation-produced aspartyl-containing peptide (left column). The mass spectra acquired across the deamidation peaks are shown in the right column.

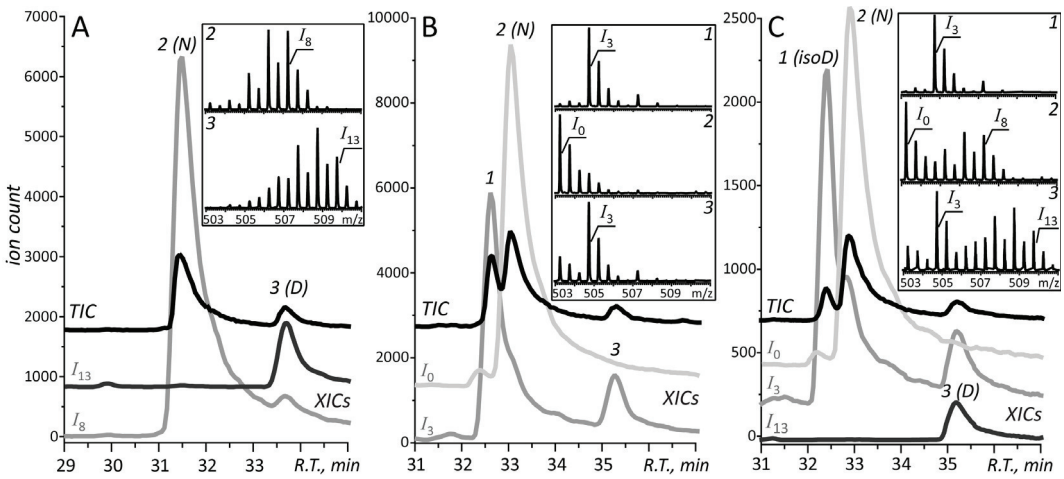


Figure 3-2. LC-MS analysis of the O18-labeled acid-catalyzed (A) and base-catalyzed (B) deamidation standards generated using model peptide EWSVNSVGK, and their mixture (C). The total ion chromatogram (TIC) is shown in black trace, and the extracted ion chromatogram (XIC) of ionic species I_0, I_3, I_8, and I_{13} are shown in grey trace, respectively. The elution peaks are labeled with numbers, and the corresponding mass spectra are shown in the insets. N: Asn; D: Asp; isoD: isoAsp.

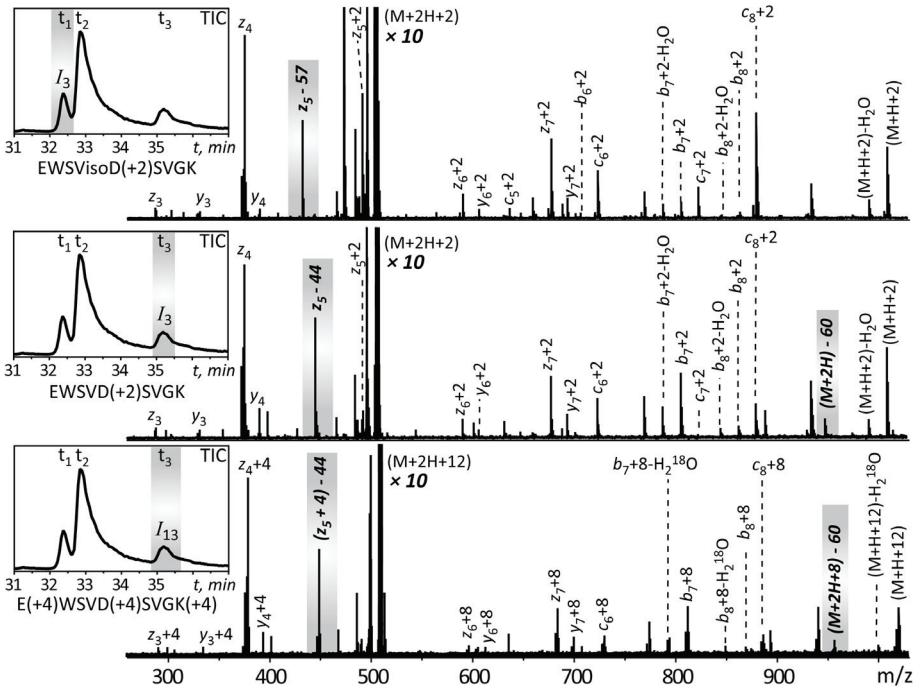


Figure 3-3. ECD MS/MS analysis of O18-labeled deamidation standards (EWSVNSVGK). The inset in each panel represents the total ion chromatogram (TIC), and the ion selected for each ECD MS/MS analysis is highlighted by the grey area. The signature ions are also highlighted by the grey area. The number in brackets after the amino acid residues indicates the mass increment in this residue due to O18 incorporation.

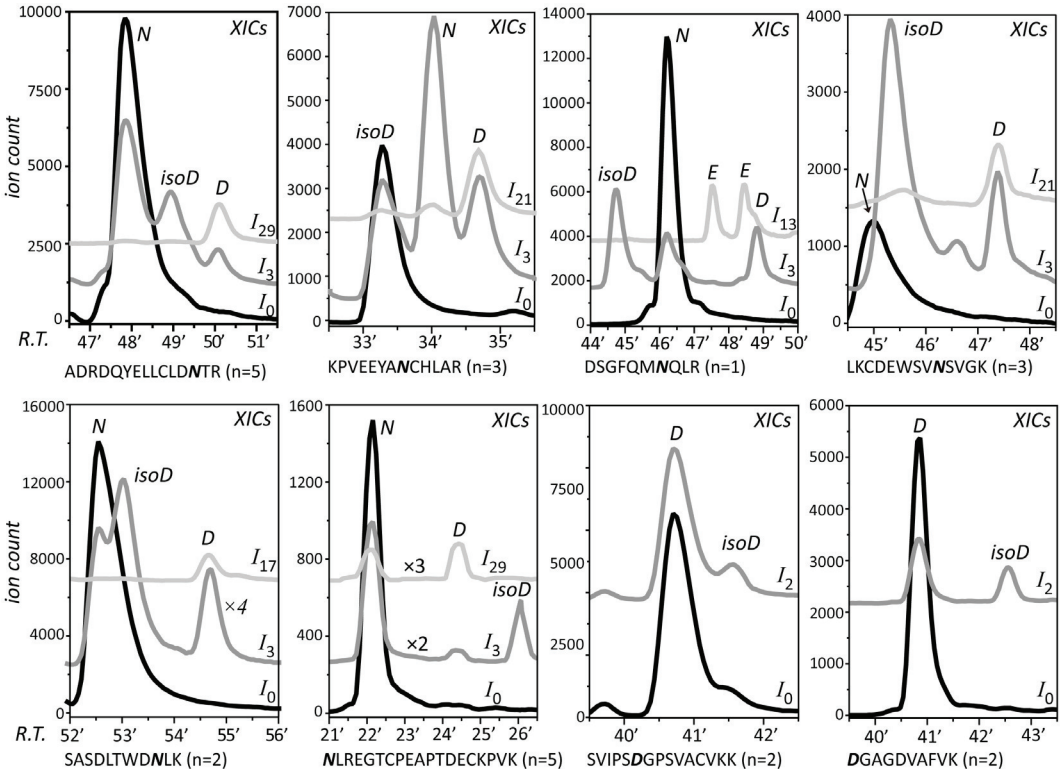


Figure 3-4. The aspartyl- and isoaspartyl-containing peptides assignment from tryptic digestion of hTf. The number in brackets after the peptide sequence indicates the number of acidic residues and carboxymethylated Cys residues in this peptide. XICs of ionic species I_0, I_3 and I_{4n+9} are shown in black, dark grey and light grey, respectively. N: Asn; D: Asp; isoD: isoAsp; E: Glu.

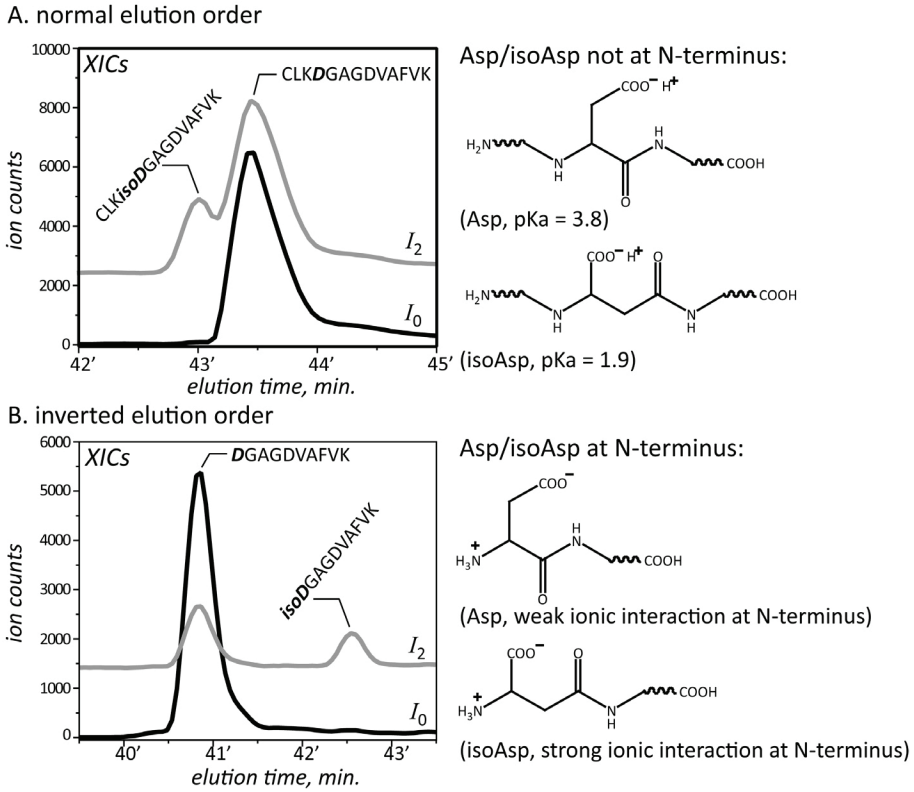


Figure 3-5. Examples of peptides exhibited both normal and inverted elution order for their aspartyl- and isoaspartyl-forms, and the possible explanations. The XICs of ionic species I_0 and I_2 are shown in black and grey, respectively.

CHAPTER 4

RANKING THE SUSCEPTIBILITY OF PROTEIN DISULFIDE BONDS TO REDUCTION

4.1 Introduction

Disulfide bonds are generally considered to play a significant role in maintaining and stabilizing structure of proteins, primarily by decreasing the entropy of the unfolded state [106]. They are frequently found in extracellular proteins where they can promote the resistance of proteins to proteolysis and denaturation in the harsh extracellular environment. In addition to the structural role, some disulfide bonds also possess roles in regulating molecular functions [107, 108]. For example, catalytic disulfide bonds present a special group of functional disulfides that can catalyze the formation, reduction, or isomerization of disulfide bonds in their protein substrate [109, 110]. More recently, it has been realized some other disulfide bonds can regulate protein function allosterically by triggering a conformational change upon their reduction [111, 112]. Those disulfides are termed 'allosteric' disulfides to distinguish them from the catalytic disulfides [108]. Indeed, disulfide bond structure is critically important to maintain protein structure, stability and biological functions, and therefore a lot of effort has been devoted to characterize this unique covalent bond in protein. A variety of analytical tools have been developed to identify the linkages of disulfide bonds, including NMR analysis [113] and Edman Degradation [114]. Particularly, MS-based methods have proved to be extremely powerful to elucidate very complicated disulfide connectivity [115-119]. In

contrast, the stability of those disulfide bonds, particularly under reductive conditions, has been less appreciated, although it can play a significant role in protein stability and function. For example, reduced binding affinity of rabbit IgG and human IgG to the first component of complement (C1q) has been observed upon partial reduction of their disulfide bonds. In order to identify those "weak spots" or active sites, it is important to develop a method for complete ranking of susceptibility of disulfide bonds upon reduction.

4.1.1 Screening for possible functional disulfide bonds

Despite the primary structural role of protein disulfide bonds, it has become clear that some disulfide bonds, particularly in cell-surface proteins, are targets of redox regulation in the immune system [120]. For example, during the priming and activation of T cells, an elevated thiol level on cell surface can be observed due to reduction of disulfide bonds in cell-surface proteins by thioredoxin [121]. Indeed, those functionally important disulfide bonds are frequently redox-labile, since they have to be sensitive to the changes in redox condition in order to switch between bonded and unbounded states. Taking advantage of this feature, a proteomic-based approach has been reported to systematically screen for membrane proteins containing labile disulfide bonds. The results suggested that reduction of certain disulfide bonds in a wide range of membrane proteins was common during immune activation [107].

4.1.2 Disulfide bonds responsible for covalent aggregation of proteins

Formation of covalent aggregates of proteins via intermolecular disulfide scrambling is a great concern for protein therapeutics, which can result in both diminished bioactivity and increased immunogenicity risk [122, 123]. For example, the dimerization of IgG2 molecule was reported to be a result of disulfide bond formation between cysteines located in the hinge region [124, 125]. Furthermore, intermolecular disulfide scrambling was also attributed to the particle formation of antibody under agitation stress [126]. Apparently, labile disulfide bonds in proteins are the primary targets for the formation of nonnative disulfide bonds, by providing active free cysteines, which emphasizes the importance of ranking the susceptibility of disulfide bonds in a protein to reduction.

4.1.3 Disulfide-based Protein Modification

Recently, a novel disulfide-based protein modification approach was reported in several studies, in which a disulfide bond is selectively reduced followed by *bis*-alkylation to insert the modification reagent (e.g., PEG) [127, 128]. Because the modification site can be well controlled by selective reduction under mild conditions, site-specific modification can be achieved; this is an attractive feature compared to traditional amine-based modification. In addition, since the reduced disulfide bond is re-bridged after *bis*-alkylation, the protein's tertiary structure is normally maintained and the biological activity is preserved [129]. Frequently, crystal structure of a protein can be

used to estimate the relative surface accessibility of each disulfide bond. More recently, a computation approach has been developed for identification of solvent-accessible disulfide bonds using published structural information [130]. However, it is important to emphasize that the static image generated from X-ray crystallography cannot fully represent the dynamic properties of a protein in solution. Furthermore, in addition to solvent accessibility, other factors (as discussed below) have also been shown to contribute to the susceptibility of disulfide bonds upon reduction. Thus, a direct measurement of susceptibility of disulfide bonds in protein is favored in order to accurately predict the disulfide-based modification site.

4.1.4 Factors contribute to susceptibility of protein disulfide bonds to reduction

Obviously, disulfide bonds at the surface of a protein are probably more susceptible to reduction, since they are accessible to either small thiols or enzymatic reducing agents. However, there is evidence that, although rare, some buried disulfides can be readily reduced with the assistance of exposed disulfides via an intra-molecular thiol-disulfide exchange mechanism [131]. In addition, dynamic features of a protein in solution can also greatly affect the stability of disulfide bonds upon reduction, and these cannot be predicted based on the static image generated from X-ray crystallography [130]. A recent bioinformatics study also suggested that labile disulfide bonds are usually found to carry relatively high dihedral strain energy that make them more easy to cleave [108]. Furthermore, the presence of a positively charged neighboring residue near a

disulfide bond can also enhance its lability by providing a favorable electrostatic environment for increased local concentration of thiolated anion [132]. Finally, more recently, mechanical force was suggested as another important factor that affects the stability of disulfide bonds upon reduction, although those studies are currently limited to engineered disulfides [133, 134].

4.1.5 Ranking the susceptibility of protein disulfide bonds to Reduction

Indeed, the presence of multiple contributing factors greatly amplifies the challenge for prediction of susceptibility of disulfide bonds upon reduction. Although X-ray crystallography is extremely useful to probe the solvent accessibility and calculate the dihedral strain energy (DSE) of a disulfide bond based on protein crystal structure. For proteins without crystal structure or the resolution is insufficient, this approach is less useful. Indeed, crystallography has been proved to be particularly useful for fast- and large-scale screening of labile disulfides in the protein database [108]. However, direct measurement is still required to achieve more reliable and accurate ranking of the susceptibility of all disulfide bonds in a protein. Fortunately, the development of mass spectrometry provides a powerful tool to address this challenging task. Particularly, proteomic-based methods have been developed to determine the cysteine oxidation status using a differential alkylation approach [135]. More recently, a similar approach was applied to rank the susceptibility of disulfide bonds in human IgG1 antibodies [136]. In this study, we proposed a new LC-MS/MS based approach for the ranking of

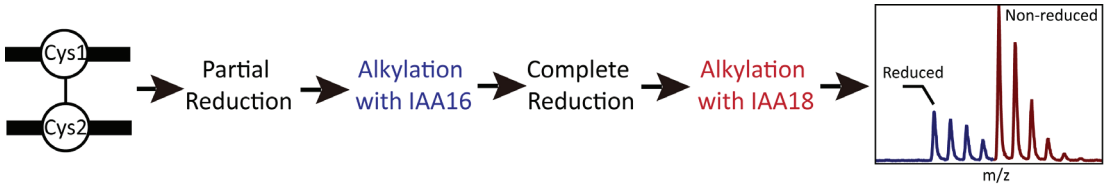
susceptibility of disulfide bonds in human Tf using O^{16}/O^{18} labeled iodoacetic acid for differential alkylation. The preparation and quantitative application of O^{18}-labeled iodoacetic acid (IAA-18) have previously been reported in our lab [137].

4.2 Results and discussion

4.2.1 Experiment workflow

15 µg of hTf was prepared in 20 µL of 50 mM potassium phosphate buffer at pH 7.4. Partial reduction was initiated by addition of DTT to a final concentration of 10 mM, and followed by incubation at 37 ℃ for 5 min. Reduction was quenched by a 25-fold dilution into the alkylation solution consisted of 10 mM of O^{16}-labeled iodoacetic acid (IAA-16) and 6 M of guanidine hydrochloride at pH 8.0. Alkylation of free cysteine residues was conducted at 37 ℃ for 30 min in the dark. Subsequently, the excess IAA-16 was removed by extensive buffer exchange into a denaturing buffer (50 mM of potassium phosphate, 6 M of guanidine hydrochloride, pH 8.0) using a Vivaspin concentrator with a 10 kDa molecular weight cut-off membrane. The complete reduction was accomplished by the addition of DTT to a concentration of 10 mM and incubation at 37 °C for 30 min. Finally, all the free cysteine residues in hTf were alkylated with 30 mM IAA-18 at 37 °C for 30 min in the dark. After tryptic digestion, peptide fragments were subjected to nanoLC-MS/MS analysis, and cysteine-containing peptides were identified and used for calculation of reduction percentage of each disulfide bond (Scheme 5).

Scheme 5



4.2.2 Determination of reduction percentage of protein disulfide bonds upon partial reduction

The IAA-18 solution can be easily prepared by exchanging carboxylic oxygen atoms of commercially available IAA in H_2O^{18} at low pH [137]. Compared to C13-labeled alkylating reagents that are more frequently used [136, 138], IAA-18 provides larger mass separation (4 Da) between the differentially alkylated peptides. Analysis of the control sample alkylated with IAA-18 solution revealed that less than 1% of alkylated peptides failed to incorporate any O^{18} atom (Figure 4-1B). In addition, it has been demonstrated that the ratio of fully labeled products (+4 Da) and partially labeled products (+2 Da) is constant among all cysteine-containing peptides [137]. This feature greatly simplifies the calculation of the reduction percentage, which can be estimated using the following equation:

$$\text{Reduction (\%)} = \frac{I_1}{I_1 + I_5 \cdot (1+\alpha)},$$

where α (α=0.14 in this work) is the fraction of partially labeled peptides; and I_1 and I_5 represent the observed relative intensities of the second isotopic peaks for O^{16}-labeled peptides and fully O^{18}-labeled peptides, respectively (Figure 4-1C). Also, control experiments with 0 minute of reduction revealed the existence of after-quench reduction (data not shown), which occurred during the alkylation step (Scheme 5). The after-quench reduction would certainly contribute to the overall reduction percentage (from 3% to 8% among all the peptides), however it can be easily subtracted during data

analysis.

4.2.3 Reduction percentage of disulfide-linked cysteine residues

An assumption was applied to the data analysis that two disulfide-linked cysteine residues should exhibit the same reduction percentage. Thus, the reduction percentage of a disulfide bond could be determined by either one of its cysteine residues. The validity of this assumption was demonstrated by six disulfides in hTf where both the cysteine residues could be measured (Figure 4-2). Indeed, identical reduction percentage (within experimental error) was observed between all the paired cysteine residues. This assumption was frequently applied in data analysis in order to improve the coverage of measurable disulfide bonds. In addition, this measurement could also be applied to assist the determination of unknown disulfide bond linkages by reducing possible combinations. For example, Cys-9 and Cys-118 in hTf were less likely to form the same disulfide bond, because they exhibited a significant difference in stability upon reduction (Figure 4-2).

4.2.4 Analysis of peptides containing two cysteine residues

In order to improve the coverage of measurable disulfide bonds, analyses of peptides with two cysteine residues were frequently required, although they produced convoluted mass spectra (Figure 4-3). To address this difficulty, the previously demonstrated assumption was applied (*see* Section 3.2.3). For example, the reduction percentage of two disulfide bonds within peptide EGTC*PEAPTDEC*KPVK could be easily

calculated using isotopic distribution of two of its disulfide-linked peptides SAGWNIPIGLLYC*DLPEPR and APNHAVVTRKDKEAC*VHK. Even in the absence of one disulfide-linked peptide (peptide APNHAVVTRKDKEAC*VHK in this case), reduction percentage of this disulfide bond could still be calculated using a mathematic method derived from two known isotopic distributions (Figure 4-3).

In the case of peptide C*STSSLLEAC*TFR, none of the disulfide-linked peptides could be detected in LC-MS/MS analysis, and therefore tandem mass spectrum generated by CAD was applied for calculation of reduction percentage. As shown in Figure 4-4, the reduction percentage of the second disulfide bond could be easily calculated using the isotopic distribution of fragment ion y_7^+. Subsequently, the reduction percentage of the first disulfide bond could be calculated applying the same strategy discussed above. However, in order to acquire reliable and quantitative data from tandem mass spectra, it was critically important to ensure that ions generated from differentially alkylated peptides (ranging from 767 to 774 m/z) would be isolated to the first quadrupole and consequently get fragmented equally efficiently. The former was believed to be valid since isotope labeling should not affect the fragmentation efficiency of peptides. However, in order to isolate all the ions between the 767 to 774 m/z range for subsequent fragmentation without bias, the mass selection window was set to low. A control experiment using very low collision energy (5 eV) confirmed the equal efficiency of mass isolation (Figure 4-4b).

4.2.5 Analysis of disulfide bonds in hTf

Applying this method, the reduction percentage of 15 disulfide bonds in hTf (19 disulfide bonds in total) could be measured with good precision after 5 minutes of reduction by DTT (10 mM, 37 °C and pH 7.4), and therefore the susceptibility of those disulfide bonds to reduction could be readily ranked. As shown in Table 5, those 15 disulfide bonds exhibited a wide range of reduction percentage, ranging from 8% to 60%. In order to correlate the susceptibility with possible contributing factors, those disulfide bonds were also analyzed for solvent accessible surface areas (SASA), secondary structure feature and dihedral strain energy (DSE) [139]. All those structural analyses were based on a recently solved crystal structure of diferric-hTf (PDB ID 8v83) [140].

Not surprisingly, solvent accessibility was found to play one of the most important roles to modulate the susceptibility of a disulfide bond to reduction. For example, buried disulfide bonds such as Csy9-Cys48, Cys19-Cys39, Cys227-241, Cys355-368, Cys450-Cys523, Cys484-Cys498 and Cys563-577 all exhibited limited reduction with less than 20% reduced in 5 minutes. In contrast, solvent accessible disulfide bonds such as Csy137-Csy331 and Cys339-Csy596 were observed to be much more susceptible to reduction. In addition, flexibility of the protein segments where a disulfide bond is located was also observed to be important. For example, disulfide bonds connecting an α-Helix and a β-strand (Csy9-Cys48, Cys19-Cys39 and Cys355-Cys368) were found to be much more resistant to reduction comparing to those with one cysteine residues located

at a flexible loop region (Csy137-Cys331 and Cys339-Cys596). Although it is difficult to evaluate the effect of flexibility independently without the interferences from other factors (e.g., disulfide bonds located in flexible region are frequently also solvent accessible), it is reasonable to speculate that good flexibility can enhance the interaction of a disulfide bond with reducing agents, and therefore increase its reductive susceptibility. Furthermore, dihedral strain energy (DSE) has been recognized as an important factor for disulfide bond stability for decades. Applying five successive χ angles, strain energy of a disulfide bond can be estimated using the following empirical formula:

$$\begin{aligned} DSE\,(kJ\,mol^{-1}) = {} & 8.37(1+\cos 3\chi_1) + 8.37(1+\cos 3\chi_1') + 4.18(1+\cos 3\chi_2) + \\ & 4.18(1+\cos 3\chi_2') + 14.64(1+\cos 2\chi_3) + 2.51(1+\cos 3\chi_3) \end{aligned}$$

where χ_1 and χ_1' are the dihedral angles of two C_α-C_β bonds, χ_2 and χ_2' are the dihedral angles of two C_β-S_γ bonds, and χ_3 is the dihedral angle of S_γ-S_γ bond [141]. For example, two disulfide bonds Cys339-Cys596 and Cys345-Cys377 in the C-lobe of hTf were calculated to carry exceptionally high strain energy of 39.5 KJ/mol and 22.7 KJ/mol respectively (a mean value of 14.8 KJ/mol was found in a recent study of 6874 unique disulfide bonds [108]). Interestingly, both the two disulfide bonds exhibited unusually high susceptibility to reduction, particularly when comparing to their homologous disulfide bonds Cys9-Cys48 and Cys137-Cys331 in the N-lobe (Table 5). Apparently, multiple factors could modulate the stability of a disulfide bond simultaneously, emphasizing the indispensability of applying a direct measurement rather than

prediction based on crystal structure alone.

4.2.6 Disulfide bonds constraining the lobe movement in hTf

As shown in Table 5, disulfide bond Cys345-Cys377 is nearly three times more susceptible to reduction than its homologous disulfide bond Cys9-Cys48 (reduction percentage of 29% and 11% respectively), despite the fact that those two disulfide bonds are both completely buried and share exactly the same nearby residues and secondary structure features. Similarly, disulfide bond Cys339-Cys596 is significantly more labile than disulfide bond Cys137-Cys331, although they are very similar in solvent accessibility and local environment. Indeed, this discrepancy in stability can be explained by the difference in dihedral strain energy of those disulfide bonds (Table 5). Crystal structure reveals that disulfide bonds Cys339-Cys596 and Cys345-Cys377 are both located within or near the linker region of hTf that connects the N-lobe and C-lobe (Figure 4-5). Interestingly, although the linker region is not structured in hTf, it adopts a helical conformation in its family member lactoferrin. In addition, the unstructured linker region in hTf possesses a unique disulfide bond Cys339-Cys596 that is missing in both lactoferrin and ovotransferrin. As a result, function of this disulfide bond was predicted to constrain the relative movement between the N- and C-lobes [142]. This assumption agrees very well with our observations. It was speculated that the tension in the linker region generated from constraining the lobe movements could contribute to the elevated strain energy of disulfide bond Cys339-Cys596, by distorting its thermally

stable configuration. Similarly, disulfide bond Cys345-Cys377 could also be destabilized by the same mechanism, since it is located right after the linker region.

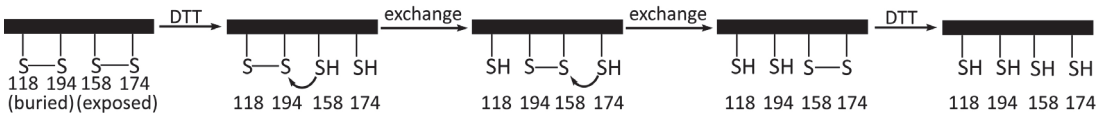
In order to demonstrate this hypothesis, reduction percentages of those two disulfides were also measured in apo hTf. It was speculated that the tension in the linker region would be released after iron removal due to the opening of both lobes, and therefore the two stressed disulfide bonds could be stabilized. Indeed, upon iron removal the reduction percentages of both disulfide bonds Cys339-Cys596 and Cys345-Cys377 were reduced by ca. 30%. In contrast, most of the other disulfide bonds maintained the reductive susceptibility after the conformational change (Figure 4-6).

4.2.7 Possible disulfide-scrambling site in hTf

Disulfide bond Cys118-Cys194 was another labile disulfide bond that can be easily overlooked by structure-based screening methods. The crystal structure of hTf suggested that this disulfide bond was completely buried and the calculated dihedral strain energy was only slightly above the mean value (Table 5). In addition, data from hydrogen/deuterium exchange MS experiments also suggested this disulfide bond was located in a rigid region, thus lacking of flexibility. However, as shown in Table 5, this disulfide bond was observed to be extremely labile upon reduction. Close examination of the crystal structure revealed a disulfide-rich segment aligning closely to this disulfide bond, particularly disulfide bond Cys158-Cys174 (Figure 4-7). Although those nearby disulfide bonds could not be detected due to the large size of tryptic fragment, they

were believed to be highly susceptible to reduction because of good solvent accessibility and flexibility. Therefore, an intra-molecular thiol-disulfide exchange mechanism was proposed to explain the lability of disulfide bond Cys118-Cys194 (Scheme 6). As a result, this region in hTf was also speculated to be a highly possible disulfide-scrambling site.

Scheme 6



4.3 Conclusion

A new LC-MS/MS based method has been developed for ranking of susceptibility of disulfide bonds in protein upon reduction using a differential alkylation strategy. Based on the isotopic distributions of cysteine-containing peptides from both MS1 and MS/MS spectra, nearly all the detectable disulfide bonds could be measured for reduction percentage upon partial reduction. Applying this method, 15 disulfide bonds in hTf were ranked based on susceptibility to reduction, and subsequently possible contributing factors such as solvent accessibility, flexibility and dihedral strain energy were discussed. Particularly, two disulfide bonds Cys339-Cys596 and Cys345-Cys377 in the linker region of hTf were demonstrated to be responsible for constraining the relative movement of N- and C-lobe. In addition, lability of another disulfide bond Cys118-Cys194 was believed to be a result of an intra-molecular thiol-disulfide exchange event, indicating a possible disulfide-scrambling site.

Table 5. Analysis of disulfide bonds in hTf

Cys-1	Cys-1 secondary structure	Cys-1 SASA (A^2)	Cys-2	Cys-2 secondary structure	Cys-2 SASA (A^2)	DSE (KJ/mol)	Reduction percentage
9	extended strand	0	48	alpha helix	0	13.0	11.2±1.3%
19	alpha helix	0	39	Strand	38	11.8	8.6±1.1%
118	strand	0	194	alpha helix	0	16.3	35.8±0.6%
227	loops or irregular	2	241	bend	1	14.6	11.6±0.7%
137	3_{10} helix	47	331	loops or irregular	44	15.7	43.8%±1.2%
339	loops or irregular	46	596	alpha helix	3.5	39.5	60.7%±5.0%
345	strand	1	377	alpha helix	0	22.7	28.6±1.7%
355	alpha helix	0	368	extended strand	19	8.38	15.5±1.6%
402	hydrogen bonded turn	16	674	alpha helix	26	5.64	21.0±1.5%
418	3-helix (3/10 helix)	11	637	extended strand	7.2	14.7	26.1±0.7%
450	strand	0	523	alpha helix	0	16.1	17.0±1.8%
474	bend	46	665	3_{10} helix	33	13.4	24.8±0.9%
484	strand	0	498	loops or irregular	6	8.52	18.3±0.5%
563	loops or irregular	6	577	bend	3	21.0	15.1±1.1%
615	loops or irregular	53	620	loops or irregular	12	14.3	32.3±3.8%

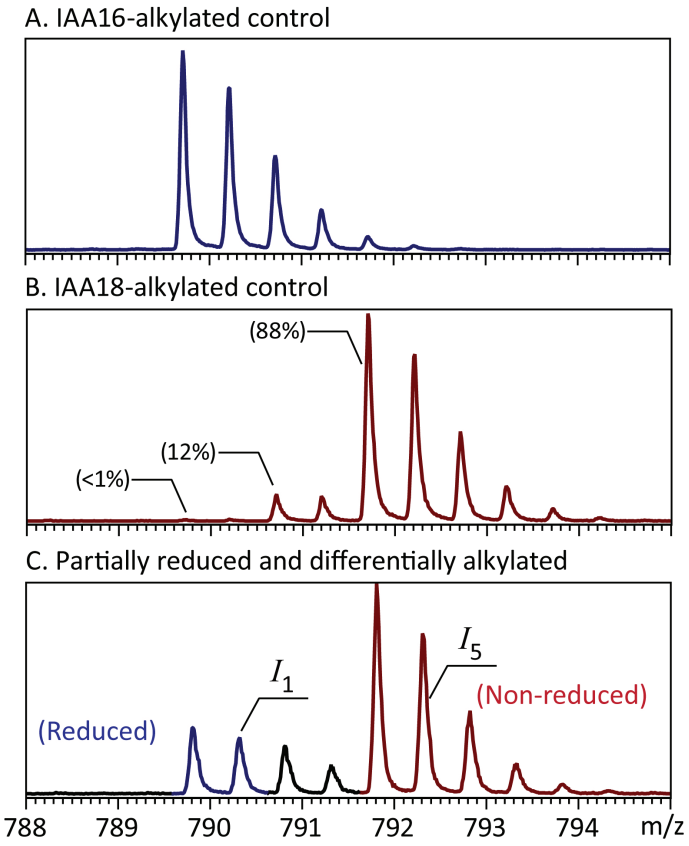


Figure 4-1. Mass spectra of a doubly charged peptide ion FDEEFSEGC*APGSK alkylated with either IAA-16 (A) or IAA-18 (B). C: Mass spectrum of differentially alkylated peptide FDEEFSEGC*APGSK after partial reduction (10 mM of DTT, 37 ℃, pH 7.4, and 5 min of incubation).

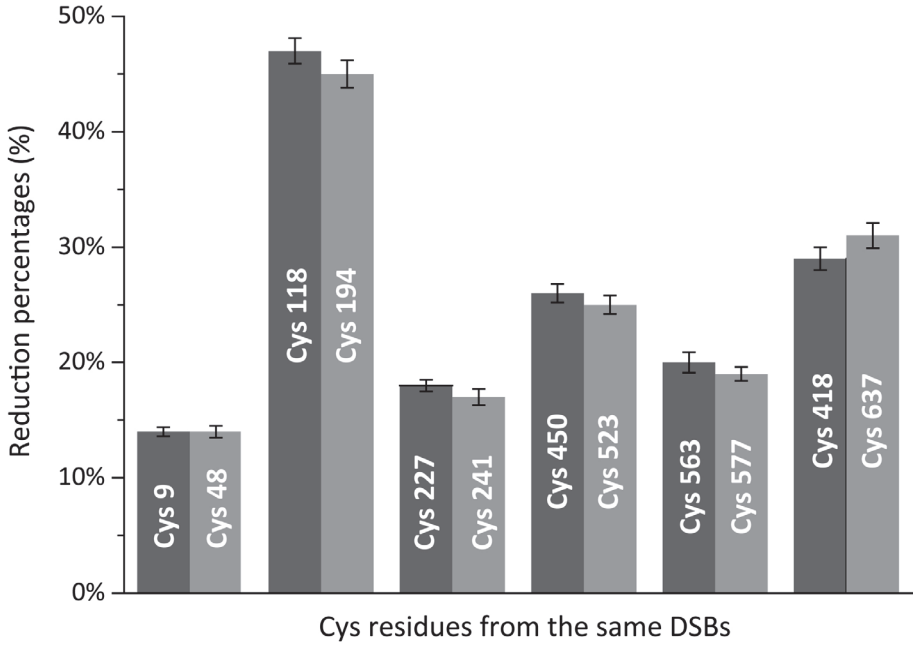


Figure 4-2. Reduction percentages of cysteine residues from the same disulfide bonds. The values are shown by different bars (dark grey and light grey) in pairs. The error bar represents the standard deviation of three replicate measurements.

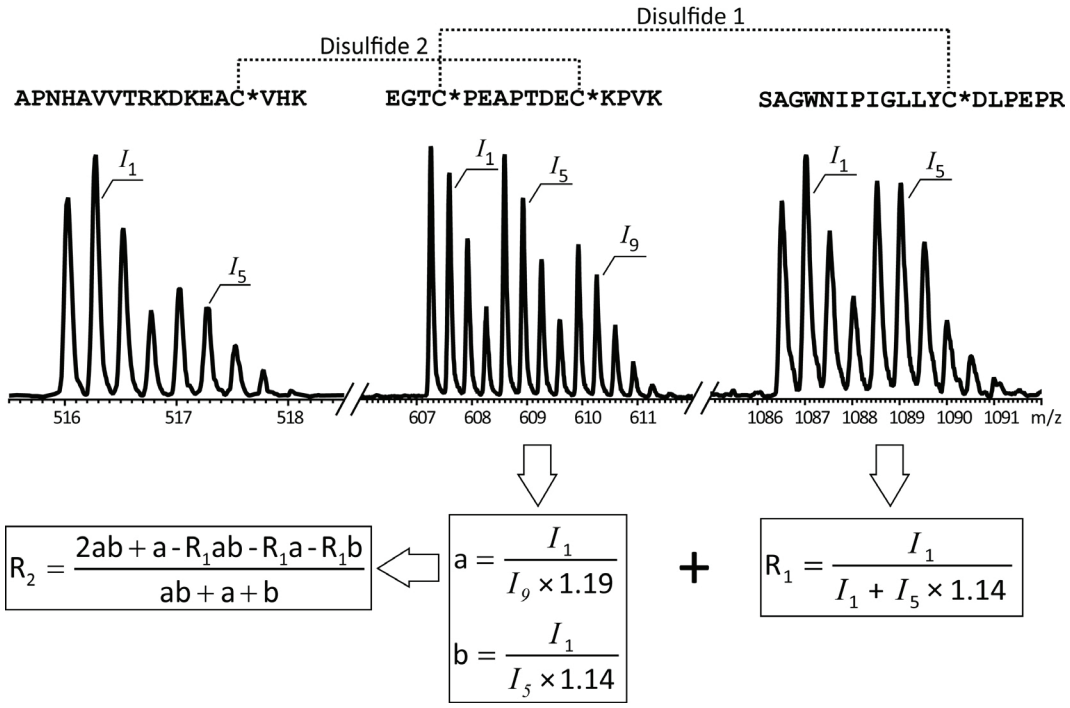


Figure 4-3. Mass spectra of three peptides that linked by two disulfide bonds. Equations on the bottom represent the calculation method for reduction percentage of one disulfide bond using the other two isotopic distributions.

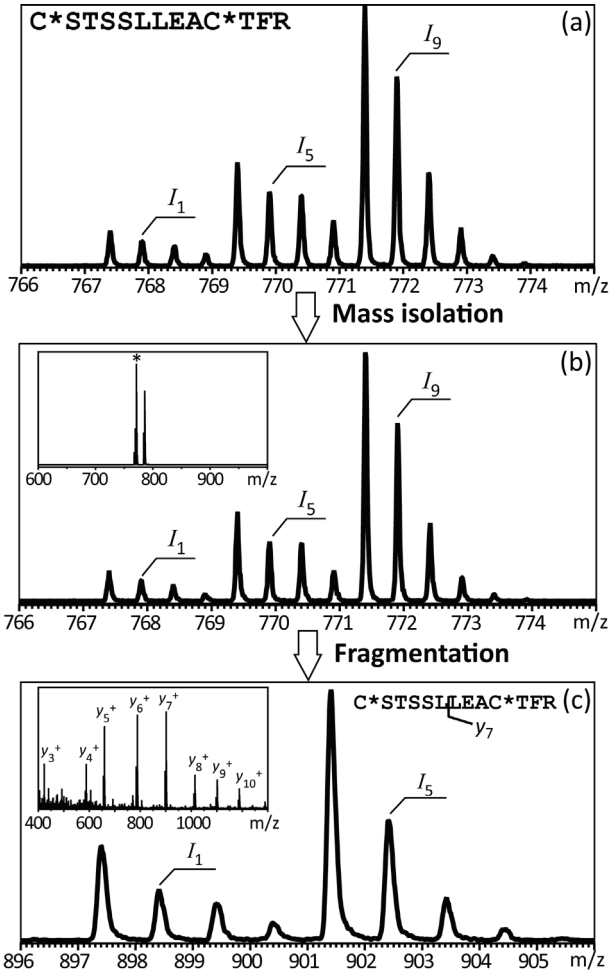


Figure 4-4. (a) Isotopic distribution of differentially alkylated peptide C*STSSLLEAC*TFR. (b) Mass isolation of the whole isotopic cluster to the first quadrupole using 'low resolution mass selection' setting and low collision energy (5 eV). (c) Isotopic distribution of fragment ion y_7^+ from differentially alkylated peptide C*STSSLLEAC*TFR.

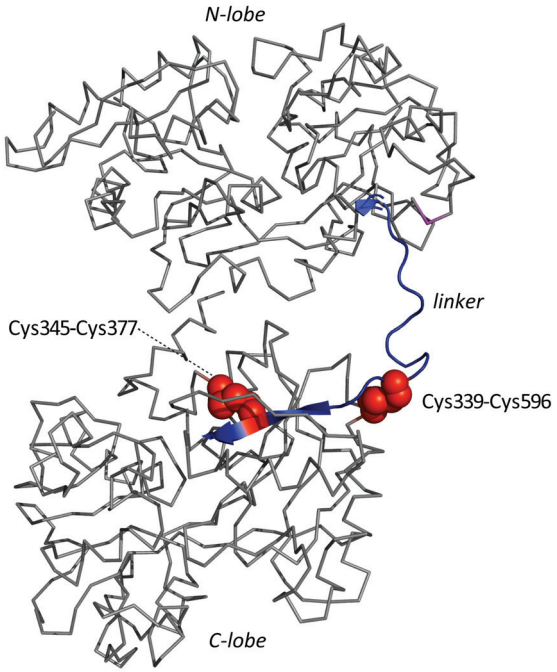


Figure 4-5. A ribbon representation of the crystal structure of holo-hTf (3v83). Two disulfide bonds with high strain energy are shown in red.

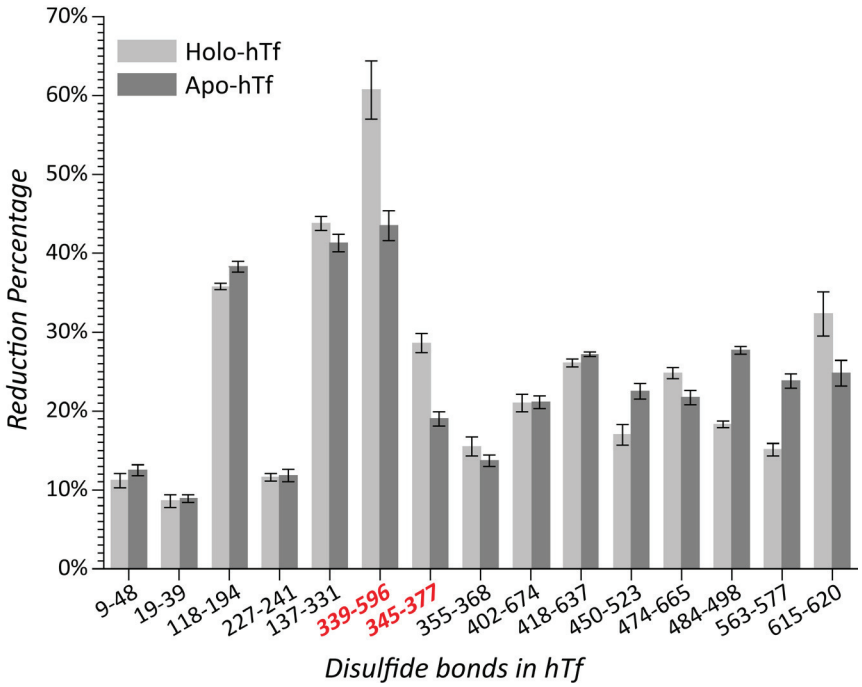


Figure 4-6. Reduction percentages of hTf disulfide bonds in both holo-form (light grey bars) and apo-form (dark grey bars). The error bar represents the standard deviation of three replicate measurements.

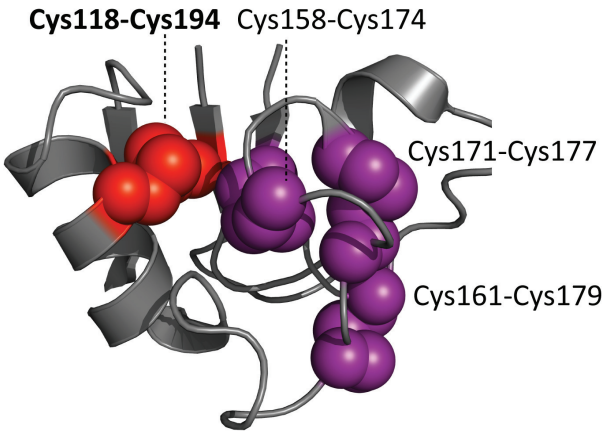


Figure 4-7. A cartoon representation of a possible disulfide-scrambling site in hTf. Disulfide bond Cys118-Cys194 is shown in red. The three nearby disulfide bonds are shown in purple.

CHAPTER 5

GROWTH HORMONE-TF FUSION PROTEIN FOR SUCCESSFUL ORAL DELIVERY

This work was done in collaboration with Dr Cedric E. Bobst in our lab. Cedric performed the SEC and native ESI MS analyses of GHT fusion protein and the binding with TfR. He also investigated the stability of GHT in a stomach-like environment. I performed all the bottom-up analyses for the identification of GHTx and the binding between GHTx and TfR.

5.1 Introduction

In spite of the great success achieved by protein drugs in treating human diseases in the last decade, the development of orally deliverable protein drugs has been notably slow [10]. Indeed, the normal fate of protein drugs after oral administration is degradation due to the harsh gastrointestinal (GI) tract environment (not only the extreme pH in stomach but also the presence of various proteases). Even the survival of such conditions cannot guarantee efficient delivery, which also requires the traversing of intestinal-epithelial barrier. One of the very few successful examples is a recently designed human growth hormone/transferrin fusion protein (GHT) that exhibited measurable therapeutic response after oral administration. In that study, significant weight gain was observed in growth hormone-deficient mice after oral administration with GHT fusion protein, but not with growth hormone alone [8]. However, molecular basis for this rare success was not elaborated, and elucidation might provide important

insights for the design of orally deliverable protein drugs in the future.

5.2 Results and discussion

Mass spectrometry has become a powerful tool that can assist all stages of protein drug development. Using MS-based approaches, protein drugs can be characterized at different levels ranging from protein identification, covalent structure, conformation, interaction with physiological partners and PK/PD studies. Thus, in this study the GHT fusion protein was analyzed primarily by MS-based methods.

5.2.1 Protein identification by native ESI MS and a bottom-up approach

Size exclusion chromatography analysis of GHT clearly suggested two distinct fractions, with one eluting very close to the column's void volume and the second one eluted slightly earlier than the Tf control. Subsequently, the species in the second fraction was identified as the monomeric form of GHT (GHT1) based on the accurate mass provided by native ESI MS analysis. However, a similar strategy could not be applied for the first fraction due to the large size of the molecular species (light scattering measurements indicated that components in this fraction were submicron particles). Instead, a so-called "bottom-up" approach [143] was applied for protein identification, in which the proteins of interest are digested by endoproteases, followed by LC-MS/MS analysis of proteolytic peptides and finally peptide mass fingerprint analysis or protein-database search to identify precursor protein(s). Briefly, protein samples were denatured in 6 M of guanidine·HCl buffer, followed by reduction and alkylation of cysteine residues, and finally digestion by trypsin. Analysis of the LC-MS data using the amino acid sequence of

GHT identified fragments derived from both Tf and growth hormone moieties. In addition, peptides from the helical linker region of GHT were also identified (Figure 5-1). This experiment unequivocally confirmed the presence of GHT in the submicron particle fraction, in which this protein existed as large soluble aggregates (GHTx).

It is also worthy to mention that during the bottom-up experiments, GHTx exhibited significant resistance to tryptic digestion in comparison to the Tf control, suggesting a promoted ability of GHTx to resist hydrolysis. Thus, it was speculated that the aggregation state of GHT might play an essential role in its ability to survive the harsh stomach environment. To validate this hypothesis, the stability of GHT1 and GHTx in the stomach-like environment (pH 3.5, 37 °C and in the presence of pepsin) was investigated by Cedric. By monitoring absorbance (214 nm) at the intact protein peak, the extent of proteolysis of both GHT1 and GHTx was measured by SEC. Indeed, GHTx displayed considerable stability in stomach-like environment, with a negligible decrease in abundance after 5 hours of exposure to acid and pepsin. In stark contrast, only 15% of GHT1 protein survived after 1.5 hours, demonstrating the essential role of protein aggregation to withstand the harsh GI tract environment.

5.2.2 Binding ability of GHT1 and GHTx to TfR receptor

Nevertheless, survival of GHTx in GI tract environment cannot guarantee effective delivery, since it still needs to cross the intestinal-epithelial barrier. Due to the wide distribution of Tf receptor (TfR) on the inner lining of GI tract [7], the Tf moiety was

believed to play a critical role in this step by taking advantage of receptor-mediated transcytosis. Thus, it was necessary to demonstrate the ability of the complex to bind to TfR. Using both SEC and ESI MS, we investigated the binding ability of GHT1 to TfR. Mixing purified GHT1 with an excess of TfR led to a new peak that eluted earlier than free GHT1, indicating formation of a complex. The identity of this complex was confirmed using a new ESI MS-based strategy developed in our lab [144]. Both complexes GHT1·TfR and $(GHT1)_2$·TfR were observed, indicating the same binding stoichiometry as Tf/TfR complexes. In addition, since excess of TfR led to complete elimination of free GHT1, the binding affinity could be estimated to be submicromolar range (all protein concentrations were in low- and submicromolar ranges).

Unfortunately, similar approaches could not be applied to investigate the binding ability of GHTx to TfR. Owing to the large size, a retention time shift of GHTx upon binding to TfR was simply not possible. However, we did note that addition of an excess of TfR to GHTx led to a significant change in area of the early-eluting peak, indicating a binding event between GHTx and TfR. In order to demonstrate the specific feature of the binding, a control experiment was conducted where addition of an excess of albumin to GHTx did not result in detectable change in area of the early-eluting peak. Although the components in the early-eluting peak could not be analyzed by MS directly due to the large size, isolation of this fraction followed by bottom-up analysis revealed the presence of both GHT and TfR.

As shown in Table 6, the top 5 protein hits by MASCOT MS/MS ion search engine

was listed. Furthermore, identification of one such tryptic fragment (LTTDFGNAEK) from TfR was demonstrated in Figure 5-2. Both MS/MS pattern (Figure 5-2C) and retention time (Figure 5-2D) of this fragment derived from the early-elution peak closely matched that derived from the TfR control sample, indicating presence of TfR in the early-elution peak. Those results suggested that the aggregate state of GHT remained binding affinity to TfR, which was essential for successful oral delivery.

5.3 Conclusion

This study investigated the molecular basis responsible for a rare success of GHT as an orally deliverable protein drug. Once again, Tf molecule was demonstrated as a promising drug delivery vehicle to overcome the intestinal-epithelial barrier. In addition, this study emphasized the important role of soluble aggregates of proteins to achieve successful oral delivery, by promoting the stability of protein drugs in harsh GI tract environment. This example argued the traditional negative impression of protein aggregation phenomena, particularly in the biopharmaceutical field. Instead, it suggested a possibility of exploiting protein aggregation in the design of efficient oral deliverable protein drugs. Finally, this study further demonstrated that MS was a powerful tool for characterization of protein drugs at various levels, such as protein identification and monitoring interaction with physiological partners as shown in this example.

Table 6. Top five proteins identified in the first elution peak

Protein	Protein Score	Peptide matches
Cationic trypsin OS=Bos taurus	836	10
Transferrin receptor protein 1 OS=Homo sapiens	94	5
Transferrin OS=Homo sapiens	89	7
Keratin, type I cytoskeletal 9 OS=Homo sapiens	85	4
Human growth hormone OS=Homo sapiens	81	2
Keratin, type II cytoskeletal 1 OS=Homo sapiens	81	5

MATGSRTSLLLAFGLLCLPWLQEGSAFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQN
PQTSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEG
*IQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF***LE**
AEAAAKEAAAKEAAAKEAAAK**ALEAEAAAK**EAAAKEAAAKEAAAK**ALE**VPDKTVRWCAVSEHEATKCQSFRDH
MKSVIPSDGPSVACVKKASYLDCIRAIAANEADAVTLDAGLVYDAYLAPNNLKPVVAEFYGSKEDPQTFYYAV
AVVKKDSGFQMNQLRGKKSCHTGLGRSAGWNIPIGLLYCDLPEPRKPLEKAVANFFSGSCAPCADGTDFPQLC
QLCPGCGCSTLNQYFGYSGAFKCLKDGAGDVAFVKHSTIFENLANKADRDQYELLCLDNTRKPVDEYKDCHLA
QVPSHTVVARSMGGKEDLIWELLNQAQEHFGKDKSKEFQLFSSPHGKDLLFKDSAHGFLKVPPRMDAKMYLGY
EYVTAIRNLREGTCPEAPTDECKPVKWCALSHHERLKCDEWSVNSVGKIECVSAETTEDCIAKIMNGEADAMS
LDGGFVYIAGKCGLVPVLAENYNKSDNCEDTPEAGYFAVAVVKKSASDLTWDNLKGKKSCHTAVGRTAGWNIP
MGLLYNKINHCRFDEFFSEGCAPGSKKDSSLCKLCMGSGLNLCEPNNKEGYYGYTGAFRCLVEKGDVAFVKHQ
TVPQNTGGKNPDPWAKNLNEKDYELLCLDGTRKPVEEYANCHLARAPNHAVVTRKDKEACVHKILRQQQHLFG
SNVTDCSGNFCLFRSETKDLLFRDDTVCLAKLHDRNTYEKYLGEEYVKAVGNLRKCSTSSLLEACTFRRP

Figure 5-1. Peptide mass fingerprint analysis of high molecule weight fraction from SEC. The sequence of growth hormone is in italics, the sequence of the linker is in bold, and the rest is the sequence of transferrin. The underlined peptides were identified within a mass tolerance of 50 ppm.

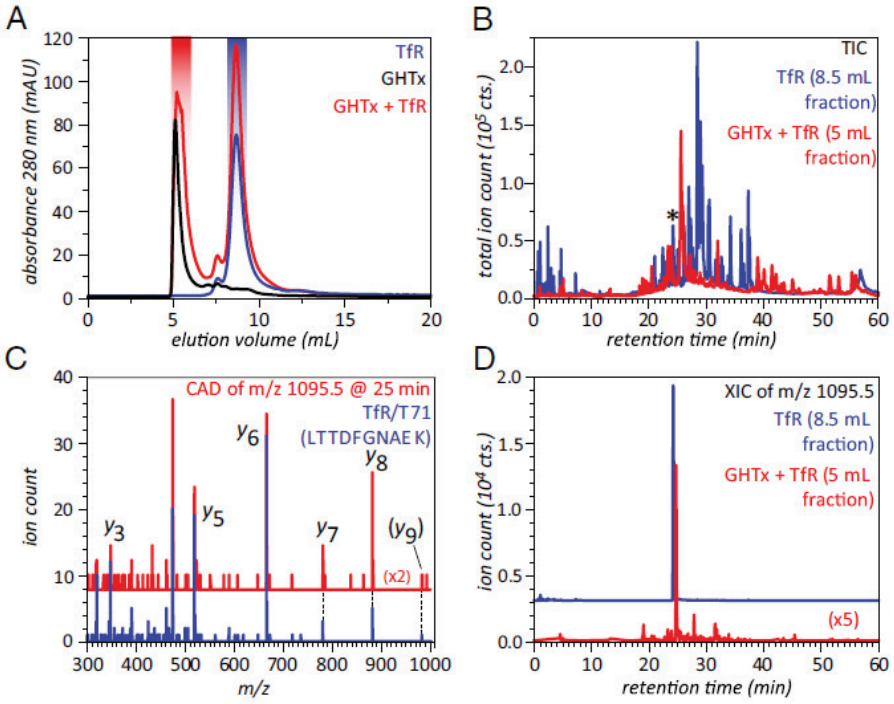


Figure 5-2. A. SEC analysis of GHTx, TfR, and their mixture. B. total ion chromatograms of tryptic digests of SEC fractions of TfR (blue trace) and GHTx/TfR mixture (red trace) highlighted in (A). C. reference MS/MS spectrum of a tryptic fragment T71 (LTTDFNAEK) of TfR derived from TfR fraction (blue) and GHTx/TfR fraction (red). D. extracted ion chromatogram of this tryptic peptide eluting at 25 min (marked with asterisk in (B). Figure is from reference [22].

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Mass spectrometry has become a powerful tool that can assist the characterization of protein drugs at different levels ranging from protein identification, covalent structure, conformation, interaction with physiological partners and PK/PD studies. In this work, we focused on the development of MS-based methods for quantitation of transferrin, a promising drug delivery vehicle, in complex biological fluids. Particularly, pitfalls in a classic approach (acid-catalyzed O^{18} labeling) to preparation of O^{18}-labeled internal standard were discovered, and a novel approach (Cys O^{18}-labeling) was developed for reliable quantitation of exogenous transferrin in biological fluids. Alternatively, a new method based on indium labeling and ICP-MS detection was developed for ultrasensitive quantitation of Tf. Following method development and validation, those methods will be applied in an animal mode study (Rat), in order to investigate the biodistribution of human Tf after intravenous injection, as well as its BBB-traversing ability. In addition, MS-based methods also provide a possibility to achieve multiplexing so that different versions of protein drugs can be analyzed simultaneously. In proteomic-based approaches, multiplexing can be achieved with relative ease using protein-specific reference peptides. However, for ICP-MS based approach, the application of multiple metal ions or isotopes as tracers will be required in order to analyze more than one type of Tf in a single experiment. The selection of suitable tracers will require extensive

analysis of the metal ion including its binding affinity to transferrin, the binding affinity of metal-coded Tf to TfR, and the ICP-MS detection of this metal.

In this work, a new method based on O^{18}-labeling and LC-MS detection was developed for unambiguous assignment of aspartyl- and isoaspartyl-containing peptides produced by Asn deamidation and Asp isomerization. This method provided a simple and reliable solution to a challenging analytical task, differentiation between deamidated isomeric peptides. In addition, this method can be easily applied to other protein drugs where deamidation is a great concern during protein production and long-term storage. Following the identification of deamidated species, accurate and sensitive quantitation of them presents another urgent and challenging task. We speculate that the application of MS-based methods in combination with stable isotope labeling will prove to be useful in this field, although elegant experimental design is required.

Furthermore, a method consisting of partial reduction, differential alkylation, and LC-MS analysis was developed in this work for ranking of susceptibility of protein disulfide bonds to reduction. Applying this method, 15 disulfide bonds in hTf was analyzed. Particularly, two disulfide bonds were found to play an important role in constraining lobe movement in hTf, which is in great agreement with other studies. In addition, a possible disulfide-scrambling site was found in hTf, although further study is required to confirm that. Furthermore, this approach will be applied to study covalent aggregates of proteins due to intermolecular disulfide scrambling. It is speculated that

the most labile disulfide bond in a protein is most likely responsible for disulfide-induced protein aggregation. This assumption will be validated by identifying the intermolecular disulfide linkages from protein aggregates induced by partial reduction.

Finally, our study also investigated the molecular basis responsible for a rare success of GHT as an orally deliverable protein drug. Tf molecule was demonstrated as a promising drug delivery vehicle to overcome the intestinal-epithelial barrier. In addition, the soluble aggregate state of GHT fusion protein was proved to be essential to achieve successful oral delivery, by promoting its stability in harsh GI tract environment. However, several questions were remained unanswered. For example, what's the nature of GHT aggregates? Will the aggregates dissociate in the circulation after crossing the intestinal-epithelial barrier? What are the differences in pharmacokinetics properties between monomeric GHT and aggregate GHT? Answering those questions will greatly benefit the design of new protein drugs suitable for oral delivery.

BIBLIOGRAPHY

1. Walsh, G., *Biopharmaceutical benchmarks 2006.* Nat Biotechnol, 2006. 24(7): p. 769-U5.
2. Wadia, J.S. and S.F. Dowdy, *Transmembrane delivery of protein and peptide drugs by TAT-mediated transduction in the treatment of cancer.* Adv Drug Deliver Rev, 2005. 57(4): p. 579-596.
3. Pardridge, W.M., *Blood-brain barrier delivery.* Drug Discov Today, 2007. 12(1-2): p. 54-61.
4. Huxley, T.H., *The Connection of the Biological Sciences with Medicine.* Science, 1881. 2(64): p. 426-9.
5. Aisen, P., *Transferrin receptor 1.* Int J Biochem Cell B, 2004. 36(11): p. 2137-2143.
6. Gatter, K.C., et al., *Transferrin Receptors in Human-Tissues - Their Distribution and Possible Clinical Relevance.* J Clin Pathol, 1983. 36(5): p. 539-545.
7. Banerjee, D., et al., *Transferrin Receptors in the Human Gastrointestinal-Tract - Relationship to Body Iron Stores.* Gastroenterology, 1986. 91(4): p. 861-869.
8. Amet, N., et al., *Human growth hormone-transferrin fusion protein for oral delivery in hypophysectomized rats.* J Control Release, 2010. 141(2): p. 177-182.
9. Morishita, M. and N.A. Peppas, *Is the oral route possible for peptide and protein drug delivery.* Drug Discov Today, 2006. 11(19-20): p. 905-910.
10. Shen, W.C., *Oral peptide and protein delivery: unfulfilled promises?* Drug Discov Today, 2003. 8(14): p. 607-608.
11. Daniels, T.R., et al., *The transferrin receptor part II: Targeted delivery of therapeutic agents into cancer cells.* Clin Immunol, 2006. 121(2): p. 159-176.
12. Hartinger, C.G., et al., *KP1019, A New Redox-Active Anticancer Agent - Preclinical Development and Results of a Clinical Phase I Study in Tumor Patients.* Chem Biodivers, 2008. 5(10): p. 2140-2155.
13. Lai, H., et al., *Effects of artemisinin-tagged holotransferrin on cancer cells.* Life Sci, 2005. 76(11): p. 1267-1279.
14. Yazdi, P.T., et al., *Influence of Cellular Trafficking on Protein-Synthesis Inhibition of Immunotoxins Directed against the Transferrin Receptor.* Cancer Res, 1995. 55(17): p. 3763-3771.
15. Raso, V. and M. Basala, *A Highly Cyto-Toxic Human Transferrin-Ricin-a Chain Conjugate Used to Select Receptor-Modified Cells.* J Biol Chem, 1984. 259(2): p. 1143-1149.
16. Weaver, M. and D.W. Laske, *Transferrin receptor ligand-targeted toxin conjugate (Tf-CRM107) for therapy of malignant gliomas.* J Neuro-Oncol, 2003. 65(1): p. 3-13.
17. Fonseca, C., et al., *Targeting of sterically stabilised pH-sensitive liposomes to human T-leukaemia cells.* Eur J Pharm Biopharm, 2005. 59(2): p. 359-366.
18. Oh, K.S., et al., *Enhancement of gene delivery to cancer cells by a retargeted adenovirus.* J Microbiol, 2005. 43(2): p. 179-182.
19. Sahoo, S.K., et al., *Efficacy of transferrin-conjugated paclitaxel-loaded nanoparticles in a murine model of prostate cancer.* Int J Cancer, 2004. 112(2): p. 335-340.
20. Lim, C.J. and W.C. Shen, *Transferrin-oligomers as potential carriers in anticancer drug delivery.* Pharm Res, 2004. 21(11): p. 1985-1992.

21. Ulbrich, K., et al., *Transferrin- and transferrin-receptor-antibody-modified nanoparticles enable drug delivery across the blood-brain barrier (BBB).* Eur J Pharm Biopharm, 2009. 71(2): p. 251-256.
22. Bobst, C.E., et al., *Mass spectrometry study of a transferrin-based protein drug reveals the key role of protein aggregation for successful oral delivery.* P Natl Acad Sci USA, 2012. 109(34): p. 13544-13548.
23. Yoon, D.J., et al., *Genetically engineering transferrin to improve its in vitro ability to deliver cytotoxins.* J Control Release, 2009. 133(3): p. 178-184.
24. Lao, B.J., et al., *Inhibition of transferrin iron release increases in vitro drug carrier efficacy.* J Control Release, 2007. 117(3): p. 403-412.
25. Harris, W.R. and L. Messori, *A comparative study of aluminum(III), gallium(III), indium(III), and thallium(III) binding to human serum transferrin.* Coordin Chem Rev, 2002. 228(2): p. 237-262.
26. Abrams, M.J. and B.A. Murrer, *Metal-Compounds in Therapy and Diagnosis.* Science, 1993. 261(5122): p. 725-730.
27. Bernstein, L.R., *Mechanisms of therapeutic activity for gallium.* Pharmacol Rev, 1998. 50(4): p. 665-682.
28. Raijmakers, P.G.H.M., et al., *Transport of Ga-67 and in-111 across a Membrane - Role of Plasma-Binding and Concentration Gradients.* Nucl Med Commun, 1992. 13(5): p. 349-356.
29. Weiss, W., et al., *Protein detection and quantitation technologies for gel-based proteome analysis.* Methods Mol Biol, 2009. 564: p. 59-82.
30. Baravalle, G., et al., *Transferrin recycling and dextran transport to lysosomes is differentially affected by bafilomycin, nocodazole, and low temperature.* Cell Tissue Res, 2005. 320(1): p. 99-113.
31. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. 389(4): p. 1017-1031.
32. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol, 2005. 1(5): p. 252-262.
33. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation.* P Natl Acad Sci USA, 1999. 96(12): p. 6591-6596.
34. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. 1(5): p. 376-386.
35. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. 17(10): p. 994-999.
36. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* P Natl Acad Sci USA, 2003. 100(12): p. 6940-6945.
37. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. 3(12): p. 1154-1169.
38. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. 75(24): p. 6843-6852.

39. Thompson, A., et al., *Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS (vol 15, pg 1895, 2003).* Anal Chem, 2003. 75(18): p. 4942-4942.
40. Yao, X.D., et al., *Proteolytic O-18 labeling for comparative proteomics: Model studies with two serotypes of adenovirus.* Anal Chem, 2001. 73(13): p. 2836-2842.
41. Sevinsky, J.R., et al., *Minimizing back exchange in O-18/O-16 quantitative proteomics experiments by incorporation of immobilized trypsin into the initial digestion step.* Anal Chem, 2007. 79(5): p. 2158-2162.
42. Staes, A., et al., *Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18.* J Proteome Res, 2004. 3(4): p. 786-791.
43. Storms, H.F., et al., *Capillary isoelectric focusing-mass spectrometry for shotgun approach in proteomics.* Electrophoresis, 2004. 25(20): p. 3461-3467.
44. Patwardhan, A.J., et al., *Quantitative proteome analysis of breast cancer cell lines using O-18-labeling and an accurate mass and time tag strategy.* Proteomics, 2006. 6(9): p. 2903-2915.
45. Eckel-Passow, J.E., et al., *Regression analysis for comparing protein samples with O-16/O-18 stable-isotope labeled mass spectrometry.* Bioinformatics, 2006. 22(22): p. 2739-2745.
46. Niles, R., et al., *Acid-Catalyzed Oxygen-18 Labeling of Peptides.* Anal Chem, 2009. 81(7): p. 2804-2809.
47. Zhang, R.J., et al., *Controlling deuterium isotope effects in comparative proteomics.* Anal Chem, 2002. 74(15): p. 3662-3669.
48. Fang, X., et al., *IgY-immunoaffinity separation and detection; Divide and conquer the proteome.* Mol Cell Proteomics, 2005. 4(8): p. S363-S363.
49. Hinerfeld, D., et al., *Serum/Plasma depletion with chicken immunoglobulin Y antibodies for proteomic analysis from multiple Mammalian species.* J Biomol Tech, 2004. 15(3): p. 184-90.
50. Ezan, E. and F. Bitsch, *Critical comparison of MS and immunoassays for the bioanalysis of therapeutic antibodies.* Bioanalysis, 2009. 1(8): p. 1375-1388.
51. Sarin, H., et al., *Effective transvascular delivery of nanoparticles across the blood-brain tumor barrier into malignant glioma cells.* J Transl Med, 2008. 6.
52. Porath, J., et al., *Metal Chelate Affinity Chromatography, a New Approach to Protein Fractionation.* Nature, 1975. 258(5536): p. 598-599.
53. Wilm, M., et al., *Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.* Nature, 1996. 379(6564): p. 466-469.
54. Ivanov, A.R., et al., *Low-attomole electrospray ionization MS and MS/MS analysis of protein tryptic digests using 20-mu m-i.d. polystyrene-divinylbenzene monolithic capillary columns.* Analytical Chemistry, 2003. 75(20): p. 5306-5316.
55. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. 20(18): p. 3551-3567.
56. Wang, S.H., et al., *Pitfalls in Protein Quantitation Using Acid-Catalyzed O(18) Labeling: Hydrolysis-Driven Deamidation.* Anal. Chem., 2011. 83(18): p. 7227-7232.
57. Borisov, O.V., et al., *Low-energy collision-induced dissociation fragmentation analysis of cysteinyl-modified peptides.* Analytical Chemistry, 2002. 74(10): p. 2284-2292.

58. Strazielle, N. and J.F. Ghersi-Egea, *Physiology of blood-brain interfaces in relation to brain disposition of small compounds and macromolecules.* Mol Pharm, 2013. 10(5): p. 1473-91.
59. Felgenhauer, K., *Protein size and cerebrospinal fluid composition.* Klin Wochenschr, 1974. 52(24): p. 1158-64.
60. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. 215(3): p. 403-10.
61. Wait, R., et al., *Proteins of rat serum urine and cerebrospinal fluid: VI. Further protein identifications and interstrain comparison.* Electrophoresis, 2001. 22(14): p. 3043-3052.
62. van Heuveln, F., et al., *Inductively coupled plasma-MS in drug development: bioanalytical aspects and applications.* Bioanalysis, 2012. 4(15): p. 1933-1965.
63. Patel, P., et al., *Isotopic labelling of peptides and isotope ratio analysis using LC-ICP-MS: a preliminary study.* Analytical and Bioanalytical Chemistry, 2008. 390(1): p. 61-65.
64. Ahrends, R., et al., *A metal-coded affinity tag approach to quantitative proteomics.* Mol Cell Proteomics, 2007. 6(11): p. 1907-1916.
65. Bunzli, J.C.G., *Benefiting from the unique properties of lanthanide ions.* Accounts Chem Res, 2006. 39(1): p. 53-61.
66. Geiger, T. and S. Clarke, *Deamidation, Isomerization, and Racemization at Asparaginyl and Aspartyl Residues in Peptides - Succinimide-Linked Reactions That Contribute to Protein-Degradation.* J Biol Chem, 1987. 262(2): p. 785-794.
67. Liu, Y.D., et al., *Human antibody Fc deamidation in vivo.* Biologicals, 2009. 37(5): p. 313-322.
68. Clarke, S., *Propensity for Spontaneous Succinimide Formation from Aspartyl and Asparaginyl Residues in Cellular Proteins.* Int J Pept Prot Res, 1987. 30(6): p. 808-821.
69. Noguchi, S., *Structural Changes Induced by the Deamidation and Isomerization of Asparagine Revealed by the Crystal Structure of Ustilago sphaerogena Ribonuclease U2B.* Biopolymers, 2010. 93(11): p. 1003-1010.
70. Charache, S., et al., *Post-Synthetic Deamidation of Hemoglobin Providence (Beta-82 Lys-]Asn Asp) and Its Effect on Oxygen-Transport.* J Clin Invest, 1977. 59(4): p. 652-658.
71. Curnis, F., et al., *Spontaneous formation of L-isoaspartate and gain of function in fibronectin.* J Biol Chem, 2006. 281(47): p. 36466-36476.
72. Friedman, A.R., et al., *Degradation of Growth-Hormone Releasing-Factor Analogs in Neutral Aqueous-Solution Is Related to Deamidation of Asparagine Residues - Replacement of Asparagine Residues by Serine Stabilizes.* Int J Pept Prot Res, 1991. 37(1): p. 14-20.
73. Huang, L.H., et al., *In vivo deamidation characterization of monoclonal antibody by LC/MS/MS.* Anal Chem, 2005. 77(5): p. 1432-1439.
74. Solstad, T. and T. Flatmark, *Microheterogeneity of recombinant human phenylalanine hydroxylase as a result of nonenzymatic deamidations of labile amide containing amino acids - Effects on catalytic and stability properties.* Eur J Biochem, 2000. 267(20): p. 6302-6310.
75. Nilsson, M.R., et al., *Low levels of asparagine deamidation can have a dramatic effect on aggregation of amyloidogenic peptides: Implications for the study of amyloid formation.* Protein Science, 2002. 11(2): p. 342-349.
76. Mamula, M.J., et al., *Isoaspartyl post-translational modification triggers autoimmune responses to self-proteins.* Journal of Biological Chemistry, 1999. 274(32): p. 22321-22327.

77. Doyle, H.A., et al., *Altered immunogenicity of isoaspartate containing proteins.* Autoimmunity, 2007. 40(2): p. 131-137.
78. Shimizu, T., et al., *Biological significance of isoaspartate and its repair system.* Biol Pharm Bull, 2005. 28(9): p. 1590-1596.
79. Aswad, D.W., et al., *Isoaspartate in peptides and proteins: formation, significance, and analysis.* J Pharmaceut Biomed, 2000. 21(6): p. 1129-1136.
80. Kaltashov, I.A., et al., *Advances and challenges in analytical characterization of biotechnology products: Mass spectrometry-based approaches to study properties and behavior of protein therapeutics.* Biotechnol Adv, 2012. 30(1): p. 210-222.
81. Edman, P., *Method for Determination of the Amino Acid Sequence in Peptides.* Acta Chem Scand, 1950. 4(2): p. 283-293.
82. Paranandi, M.V., et al., *Deamidation and Isoaspartate Formation during in-Vitro Aging of Recombinant Tissue-Plasminogen Activator.* J Biol Chem, 1994. 269(1): p. 243-253.
83. Schurter, B.T. and D.W. Aswad, *Analysis of isoaspartate in peptides and proteins without the use of radioisotopes.* Anal Biochem, 2000. 282(2): p. 227-231.
84. Johnson, B.A. and D.W. Aswad, *Optimal Conditions for the Use of Protein L-Isoaspartyl Methyltransferase in Assessing the Isoaspartate Content of Peptides and Proteins.* Analytical Biochemistry, 1991. 192(2): p. 384-391.
85. Kameoka, D., et al., *A method for the detection of asparagine deamidation and aspartate isomerization of proteins by MALDI/TOF-mass spectrometry using endoproteinase Asp-N.* J Biochem, 2003. 134(1): p. 129-135.
86. Ni, W.Q., et al., *Analysis of Isoaspartic Acid by Selective Proteolysis with Asp-N and Electron Transfer Dissociation Mass Spectrometry.* Anal Chem, 2010. 82(17): p. 7485-7491.
87. Winter, D., et al., *Separation of peptide isomers and conformers by ultra performance liquid chromatography.* J Sep Sc, 2009. 32(8): p. 1111-1119.
88. Krokhin, O.V., et al., *Deamidation of -Asn-Gly-sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis.* Anal Chem, 2006. 78(18): p. 6645-6650.
89. Sargaeva, N.P., et al., *Sequence-specific predictive chromatography to assist mass spectrometric analysis of asparagine deamidation and aspartate isomerization in peptides.* Electrophoresis, 2011. 32(15): p. 1962-1969.
90. Harris, R.J., et al., *Identification of multiple sources of charge heterogeneity in a recombinant antibody.* J Chromatogr B, 2001. 752(2): p. 233-245.
91. Kossiakoff, A.A., *Tertiary Structure Is a Principal Determinant to Protein Deamidation.* Science, 1988. 240(4849): p. 191-194.
92. Gonzalez, L.J., et al., *Differentiating alpha- and beta-aspartic acids by electrospray ionization and low-energy tandem mass spectrometry.* Rapid Commun Mass Sp, 2000. 14(22): p. 2092-2102.
93. Lehmann, W.D., et al., *Analysis of isoaspartate in peptides by electrospray tandem mass spectrometry.* Protein Science, 2000. 9(11): p. 2260-2268.

94. Cournoyer, J.J., et al., *Deamidation: Differentiation of aspartyl from isoaspartyl products in peptides by electron capture dissociation.* Protein Sci, 2005. 14(2): p. 452-463.
95. Sargaeva, N.P., et al., *Identification of Aspartic and Isoaspartic Acid Residues in Amyloid beta Peptides, Including A beta 1-42, Using Electron-Ion Reactions.* Anal Chem, 2009. 81(23): p. 9778-9786.
96. Cournoyer, J.J., et al., *Detecting deamidation products in proteins by electron capture dissociation.* Anal Chem, 2006. 78(4): p. 1264-1271.
97. Yang, H.Q., et al., *Toward Proteome-Scale Identification and Quantification of Isoaspartyl Residues in Biological Samples.* Journal of Proteome Research, 2009. 8(10): p. 4615-4621.
98. Yang, H.Q., et al., *Alzheimer's Disease and Mild Cognitive Impairment are Associated with Elevated Levels of Isoaspartyl Residues in Blood Plasma Proteins.* J Alzheimers Dis, 2011. 27(1): p. 113-118.
99. Xiao, G., et al., *O-18 labeling method for identification and quantification of succinimide in proteins.* Anal Chem, 2007. 79(7): p. 2714-2721.
100. Terashima, I., et al., *Identification of deamidation and isomerization sites on pharmaceutical recombinant antibody using (H2O)-O-18.* Anal Biochem, 2007. 368(1): p. 49-60.
101. Li, X.J., et al., *Use of O-18 labels to monitor deamidation during protein and peptide sample processing.* J Am Soc Mass Spectr, 2008. 19(6): p. 855-864.
102. Gaza-Bulseco, G., et al., *Method to Differentiate Asn Deamidation That Occurred Prior to and during Sample Preparation of a Monoclonal Antibody.* Anal Chem, 2008. 80(24): p. 9491-9498.
103. I. A. Kaltashov, C.E.B., M. Zhang, R. Leverence, D. R. Gumerov., *Advances and challenges in analytical characterization of biotechnology products: Mass spectrometry-based approaches to study properties and behavior of protein therapeutics.* Biochim. Biophys. Acta, 2012. 1820: p. 417-426.
104. Robinson, N.E. and A.B. Robinson, *Molecular clocks.* Proceedings of the National Academy of Sciences of the United States of America, 2001. 98(3): p. 944-949.
105. Niles, R., et al., *Acid-Catalyzed Oxygen-18 Labeling of Peptides.* Anal. Chem., 2009. 81(7): p. 2804-2809.
106. Betz, S.F., *Disulfide Bonds and the Stability of Globular-Proteins.* Protein Sci, 1993. 2(10): p. 1551-1558.
107. Metcalfe, C., et al., *Labile disulfide bonds are common at the leucocyte cell surface.* Open Biol, 2011. 1.
108. Schmidt, B., et al., *Allosteric disulfide bonds.* Biochemistry-Us, 2006. 45(24): p. 7429-7433.
109. Nakamura, H., *Thioredoxin and its related molecules: Update 2005.* Antioxid Redox Sign, 2005. 7(5-6): p. 823-828.
110. Holmgren, A., *Thioredoxin and Glutaredoxin Systems.* Journal of Biological Chemistry, 1989. 264(24): p. 13963-13966.
111. Chigaev, A., et al., *Conformational regulation of alpha(4)beta(1)-integrin affinity by reducing agents - "Inside-out" signaling is independent of and additive to reduction-regulated integrin activation.* J Biol Chem, 2004. 279(31): p. 32435-32443.

112. Jain, S., et al., *Role of Thiol/Disulfide Exchange in Newcastle Disease Virus Entry.* J Virol, 2009. 83(1): p. 241-249.
113. Collins, E.S., et al., *Characterisation of disulfide-bond dynamics in non-native states of lysozyme and its disulfide deletion mutants by NMR.* Chembiochem, 2005. 6(9): p. 1619-1627.
114. Strosberg, A.D., et al., *The interdomain disulfide bond of a homogeneous rabbit pneumococcal antibody light chain.* J Immunol, 1975. 115(5): p. 1422-4.
115. Chelius, D., et al., *Reversed-phase liquid chromatography in-line with negative ionization electrospray mass spectrometry for the characterization of the disulfide-linkages of an immunoglobulin gamma antibody.* J Am Soc Mass Spectr, 2006. 17(11): p. 1590-1598.
116. Zhang, M.X. and I.A. Kaltashov, *Mapping of protein disulfide bonds using negative ion fragmentation with a broadband precursor selection.* Anal Chem, 2006. 78(14): p. 4820-4829.
117. Zubarev, R.A., et al., *Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity.* J Am Chem Soc, 1999. 121(12): p. 2857-2862.
118. Wu, S.L., et al., *Mass Spectrometric Determination of Disulfide Linkages in Recombinant Therapeutic Proteins Using Online LC-MS with Electron-Transfer Dissociation.* Anal Chem, 2009. 81(1): p. 112-122.
119. Gorman, J.J., et al., *Protein disulfide bond determination by mass spectrometry.* Mass Spectrom Rev, 2002. 21(3): p. 183-216.
120. Sahaf, B., et al., *Lymphocyte surface thiol levels.* P Natl Acad Sci USA, 2003. 100(7): p. 4001-4005.
121. Lawrence, D.A., et al., *Surface thiols of human lymphocytes and their changes after in vitro and in vivo activation.* J Leukocyte Biol, 1996. 60(5): p. 611-618.
122. Jenkins, N., et al., *Post-translational modifications of recombinant proteins: Significance for biopharmaceuticals.* Mol Biotechnol, 2008. 39(2): p. 113-118.
123. Rosenberg, A.S., *Effects of protein aggregates: An immunologic perspective.* Aaps Journal, 2006. 8(3): p. E501-E507.
124. Yoo, E.M., et al., *Human IgG2 can form covalent dimers.* Journal of Immunology, 2003. 170(6): p. 3134-3138.
125. Van Buren, N., et al., *Elucidation of Two Major Aggregation Pathways in an IgG2 Antibody.* Journal of Pharmaceutical Sciences, 2009. 98(9): p. 3013-3030.
126. Brych, S.R., et al., *Characterization of antibody aggregation: Role of buried, unpaired cysteines in particle formation.* Abstr Pap Am Chem S, 2010. 239.
127. Brocchini, S., et al., *Disulfide bridge based PEGylation of proteins.* Adv Drug Deliver Rev, 2008. 60(1): p. 3-12.
128. Jones, M.W., et al., *Polymeric Dibromomaleimides As Extremely Efficient Disulfide Bridging Bioconjugation and Pegylation Agents.* J Am Chem Soc, 2012. 134(3): p. 1847-1852.
129. Shaunak, S., et al., *Site-specific PEGylation of native disulfide bonds in therapeutic proteins.* Nat Chem Biol, 2006. 2(6): p. 312-313.
130. Zloh, M., et al., *Identification and insertion of 3-carbon bridges in protein disulfide bonds: a computational approach.* Nature Protocols, 2007. 2(5): p. 1070-1083.

131. Messens, J., et al., *How thioredoxin can reduce a buried disulphide bond.* Journal of Molecular Biology, 2004. 339(3): p. 527-537.
132. Holmgren, A., *Thioredoxin.* Annu Rev Biochem, 1985. 54: p. 237-271.
133. Wiita, A.P., et al., *Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques.* P Natl Acad Sci USA, 2006. 103(19): p. 7222-7227.
134. Ainavarapu, R.K., et al., *A single-molecule assay to directly identify solvent-accessible disulfide bonds and probe their effect on protein folding.* J Am Chem Soc, 2008. 130(2): p. 436-+.
135. Schilling, B., et al., *Determining cysteine oxidation status using differential alkylation.* Int J Mass Spectrom, 2004. 236(1-3): p. 117-127.
136. Liu, H.C., et al., *Ranking the Susceptibility of Disulfide Bonds in Human IgGs1 Antibodies by Reduction, Differential Alkylation, and LC-MS Analysis.* Anal Chem, 2010. 82(12): p. 5219-5226.
137. Wang, S.H. and I.A. Kaltashov, *A New Strategy of Using O-18-Labeled Iodoacetic Acid for Mass Spectrometry-Based Protein Quantitation.* J Am Soc Mass Spectr, 2012. 23(7): p. 1293-1297.
138. Schilling, B., et al., *Determining cysteine oxidation status using differential alkylation.* Int J Mass Spectrom, 2004. 236(1-3): p. 117-127.
139. WONG, J.W.H. and P.J. HOGG, *Analysis of disulfide bonds in protein structures.* J Thromb Haemost, 2010. 8: p. 2345.
140. Noinaj, N., et al., *Structural basis for iron piracy by pathogenic Neisseria.* Nature, 2012. 483(7387): p. 53-U92.
141. Katz, B.A. and A. Kossiakoff, *The Crystallographically Determined Structures of Atypical Strained Disulfides Engineered into Subtilisin.* J Biol Chem, 1986. 261(33): p. 5480-5485.
142. Wally, J., et al., *The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding.* J Biol Chem, 2006. 281(34): p. 24934-24944.
143. Demirev, P.A. and C. Fenselau, *Mass Spectrometry for Rapid Characterization of Microorganisms.* Annu Rev Anal Chem, 2008. 1: p. 71-93.
144. Leverence, R., et al., *Noncanonical interactions between serum transferrin and transferrin receptor evaluated with electrospray ionization mass spectrometry.* P Natl Acad Sci USA, 2010. 107(18): p. 8123-8128.